

Cognitive Status and *ra*-Marked Referents of Nominal Expressions in Persian Discourse

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mahmoud Sadrai

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Professor Jeanette Gundel

May 2014

Acknowledgements

My most heartfelt thanks to four of my greatest teachers and mentors, who have had immense influence in whom I have become, personally and professionally, over the course of my studies at the University of Minnesota.

Foremost, I extend my deepest gratitude to Professor Jeanette Gundel, my supervisor for my MA thesis and my PhD dissertation. Professor Gundel's selfless time, patience, guidance, and her belief in me and my work were sometimes all that kept me going. I was not only fortunate enough to be her student, but to work with her with the Givenness Hierarchy, a framework she helped formulate. Without Professor Gundel's continuous help, support, and encouragements, be it in our weekly meetings or otherwise, this dissertation would never have been written. Thank you, Jeanette, for the countless hours you patiently spent with me in your office, for making the process so pleasant and seamless, and for the stories and the laughs we shared in the process.

In addition, I extend my deepest appreciation to my committee members, Bruce Downing, Nancy Stenson, and Iraj Bashiri, who also helped me with my MA thesis and PhD dissertation.

Professor Emeritus Bruce Downing, an amazing teacher and administrator with great passion for people and immense knowledge, was instrumental in helping me shape my views on bilingualism, language policy and planning, and translating and interpreting. His convincing me of the fact that I am, indeed, a bilingual, was not the least of his challenges working with me. Through my teaching and administrative graduate assistantship positions, and a teaching mentorship with him, I learned how a great teacher and an effective administrator works. I miss not being able to run down to Bruce's office, sometimes several times a day, to ask for his guidance. Often, when I need to think clearly and objectively, be it in my personal or professional life, I think to myself, WWBD (What Would Bruce Do?). Thank you, Bruce, for everything!

Professor Emeritus Nancy Stenson was also instrumental in helping me get to where I am today. In classrooms, as well as my frequent visits to her office, sometimes a couple times a day, I learned a great deal from Professor Stenson, including approaches to learning and teaching languages, bilingualism, and codeswitching. Professor Stenson's enthusiasm for Language and linguistics was contagious and inspirational. As I went through the dissertation writing process, I often thought about her stories of writing at the bottom of the Grand Canyon! Her being an inspiration was not the least of what helped me through the long process of writing this dissertation. I distinctly remember a moment, years ago, during one of our many discussions in which I realized what I was learning was more than just facts: I was learning how to distinguish facts from conjecture in linguistics. Thank you, Nancy, for helping me understand what constitutes evidence for claims in linguistics.

Professor Iraj Bashiri has been an inspiration and a role model for me for decades. My first course with Professor Bashiri was as an undergraduate in the Fall of 1981. I went on to take many courses from Dr. Bashiri, to the extent that I changed my undergraduate major as a result. The 1980's decade was a difficult time for students like me who were away from our families in Iran. Every semester I looked forward to taking courses from Professor Bashiri on Persian literature, poetry, and history. Thank you, Iraj, for showing me the road to becoming a life-long learner by setting an amazing example.

Getting a PhD has truly been my greatest dream in life. My American dad, Rev. Donald Jacobson once told me that it takes three generations to raise a doctor. Clearly there have been a great many folks who have helped me get to where I am today. I could not possibly begin to thank them all. However, above all, Professor Gundel, Professor Downing, Professor Stenson, and Professor Bashiri were most directly instrumental in helping me realize this dream! I am forever thankful and indebted to the four of them, not just for their help with the Ph.D. process, but for all that I learned from them over the years both in their classrooms and outside class, be it in their offices, walking down the hallways, or anywhere else we had a chance to talk. What a great blessing it has been for me to have had these folks as my teachers and mentors!

Dedication

To my mother, Pari, and to my late father, Morteza: Memories. And a dream.

“I can no other answer make but thanks,
And thanks; and ever thanks”

Abstract

This dissertation examines the referential status of the morpheme *ra* in Persian discourse.

A nominal expression in Persian can occur with or without *ra* as illustrated below:

- a.
- | | | | |
|-------|----|--------|------|
| nævar | ra | zæbt | kærd |
| tape | ra | record | did |
- 's/he recorded the tape'

But not

* 's/he recorded a tape'

- b. nævar zæbt kærd
tape record did
 ‘s/he recorded a tape’

But not

* 's/he recorded the tape'

- a. nævar zæbt shod
 tape record became
 two possible interpretations:
 ‘the tape was recorded’, or
 ‘a tape was recorded’

But not
b. * nævar **ra** zæbt shod

However, the exact nature of the conditions under which *ra* occurs is a source of disagreement in the literature on this morpheme. Various, and sometimes contradictory, assertions about the functions of this Persian morpheme have been made in the literature. While most researchers agree that 'ra' primarily marks direct objects (albeit there is evidence that it can also potentially mark other non-subjects), they differ in whether these nominal expressions are marked for definiteness (Phillott 1919, Lambton 1953, Ghomeshi 1997), or specificity (Browne 1970, Karimi 1989, Windfuhr 1990). In defining the referential status of *ra*-marked nominal expressions, I use a framework, the Givenness Hierarchy (Gundel, Hedberg, and Zacharski (1993)), that makes it possible to provide a more precise definition of 'definiteness' and 'specificity'. This framework allows for more precise and fine-grained primitive notions than the traditional concepts of specificity and definiteness. Distinctions in the Givenness Hierarchy are made in terms of

cognitive statuses (the location of referents in the memory and attention of a hearer) at the point in the discourse just before a particular linguistic form is used.

In this dissertation I examine naturally occurring data from forty (40) different sources, as well as constructed examples. The purpose for choosing a variety of sources of data for this study is to ensure that the results reflect language in use in a variety of genres and contexts. As such, I examine data that includes short stories, newspaper articles, children's stories, and a telephone conversation. The data examined contains a total of 17,611 words, yielding 408 instances of *ra*-marked DPs. Each expression containing '*ra*' is examined and coded for the highest cognitive status of the referent of the nominal expression. Using the cognitive statuses outlined in the Givenness Hierarchy, I examine the referent of each nominal expression containing '*ra*' to determine if it is type identifiable, referential, uniquely identifiable, familiar, activated, or in focus. I conclude that the referents of DPs marked with only *ra* are uniquely identifiable or higher. However, a more nuanced situation arises when *ra* co-occurs with another referential marker, namely *-i* (whose referential status remains unclear).

Table of Contents

TABLE OF CONTENTS.....	VI
LIST OF TABLES.....	VIII
LIST OF FIGURES.....	X
ABBREVIATIONS.....	XI
1 INTRODUCTION.....	1
1.1 BACKGROUND.....	10
1.1.1 <i>History of Persian Language</i>	11
1.1.2 <i>Difficulties Investigating Persian Grammar: Would the “Real” Persian Please Stand Up?</i>	21
1.2 MODERN PERSIAN IN THE CONTEXT OF THE PRESENT STUDY.....	34
2 PERSIAN GRAMMAR: A SKETCH.....	36
2.1 THE MORPHEME RA.....	43
2.1.1 <i>Direct Objects and ra</i>	48
2.1.2 <i>Transitivity, ra, and Ambiguity</i>	51
2.1.3 <i>ra and its Referential Status</i>	53
3 THE STUDY.....	55
3.1 COGNITIVE STATUSES OF RA-MARKED EXPRESSIONS IN DISCOURSE.....	55
3.1.1 <i>Definiteness, Specificity, and ra</i>	55
3.1.2 <i>The Givenness Hierarchy and the Concepts of Definiteness and Specificity</i>	63
3.2 METHODOLOGY.....	68
3.2.1 <i>Data Selection</i>	70
3.2.2 <i>Procedure for Determining Cognitive Status of ra-marked DPs</i>	78
4 RESULTS.....	82
4.1 SHORT STORIES: PROLOGUE.....	84
4.2 SHORT STORIES: INTRODUCTION.....	88
4.3 SHORT STORIES—‘SWEET HAT’ AND ‘VICTORIOUS ROMAN’.....	91
4.4 CHILDREN’S STORIES.....	96
4.5 NEWSPAPER ARTICLES: ‘SPORTS’ SECTION.....	100
4.6 NEWSPAPER ARTICLES: ‘WORLD’ SECTION.....	103
4.7 NEWSPAPER ARTICLES: ‘CULTURE’ SECTION.....	110
4.8 SPEECH DATA.....	114
4.9 SUMMARY.....	119
4.9.1 <i>The Cognitive Status of ra-Marked DPs</i>	120
4.9.2 <i>DP’s with both –i and ra</i>	123
5 CONCLUSION.....	132
5.1 SUMMARY.....	132
5.2 FUTURE RESEARCH.....	135
5.2.1 <i>The Referential Status of DPs Marked with both –i and ra</i>	135
5.2.2 <i>Definite DPs not marked with ra</i>	140
5.3 THE BIGGER PICTURE.....	142
BIBLIOGRAPHY.....	148

APPENDICES	151
APPENDIX A	152
APPENDIX B	157
APPENDIX C	162

List of Tables

Table 1: Timeline of Iranian Languages Intersecting with Persian.....	13
Table 2 Cognitive Statuses and Definitions, provided by Jeanette Gundel in personal communications	63
Table 3 Sources of Data Collected	68
Table 4 Total Number of Words in the Data and Number of Occurrences of <i>ra</i>	71
Table 5 Eight Articles from the Sports Section of 'Iran' Newspaper.....	72
Table 6 Three Articles from 'Culture' Sections of IRNA and Aftab-e Yazd.....	73
Table 7 One Article from 'World' Section of IRNA	73
Table 8 Twelve Articles from 'World' Section in <i>Keyhan</i>	74
Table 9 Six Articles from 'World' Section in <i>Hamshahri</i>	75
Table 10 Five Children's Stories	76
Table 11 Cognitive Statuses and Definitions, provided by Jeanette Gundel in personal communications (2010)	78
Table 12 Guidelines for determining cognitive status, provided by Jeanette Gundel	79
Table 13 Summary Results of Highest Cognitive Statuses of <i>ra</i> -Marked DPs in the Entire Corpus Examined.....	83
Table 14 Highest Cognitive Statuses of <i>ra</i> -Marked DPs in Prologue	85
Table 15 Highest Cognitive Statuses of <i>ra</i> -Marked DPs in Introduction	88
Table 16 Number of words and <i>ra</i> -marked DPs in Sweet Hat and Victorious Roman.....	91
Table 17 Highest Cognitive Statuses of <i>ra</i> -Marked DPs in 'Sweet Hat' and 'Victorious Roman'	91
Table 18 Number of Words and <i>ra</i> -Marked DPs in Children's Stories.....	96
Table 19 Highest Cognitive Statuses of <i>ra</i> -marked DPs in Children's Stories	96
Table 20 Eight Articles from the Sports Section of 'Iran' Newspaper.....	100
Table 21 Highest Cognitive Statuses of <i>ra</i> -Marked DPs in Sports Section.....	100
Table 22 One Article from 'World' Section of IRNA	103
Table 23 Twelve Articles from 'World' Section in <i>Keyhan</i>	104
Table 24 Six Articles from 'World' Section in <i>Hamshahri</i>	105
Table 25 Highest Cognitive Statuses of <i>ra</i> -Marked DPs in World Section.....	106
Table 26 Three Articles from 'Culture' Sections of IRNA and Aftab-e Yazd.....	110
Table 27 Highest Cognitive Statuses of <i>ra</i> -Marked DPs in 'Culture' Sections.....	110
Table 28 Highest Cognitive Statuses of <i>ra</i> -marked DP's in Speech Data.....	114
Table 29 Summary Results of Highest Cognitive Statuses of <i>ra</i> -Marked DPs in the Entire Corpus Examined.....	121
Table 30 Summary Percentage.....	122
Table 31 Summary of Highest Cognitive Statuses marked by <i>ra</i> in the Data.....	133
Table 32 'Short Stories' Results Summary	157

Table 33 Eight Articles from the Sports Section of 'Iran' Newspaper.....	157
Table 34 Three Articles from 'Culture' Sections of IRNA and Aftab-e Yazd	158
Table 35 One Article from 'World' Section of IRNA	158
Table 36 Twelve Articles from 'World' Section in <i>Keyhan</i>	159
Table 37 Six Articles from 'World' Section in <i>Hamshahri</i>	160

List of Figures

Figure 1 Some of the Major Proto-Indo-European (Circa 5000 B.C.) Languages.....	12
Figure 2 Map of Iranian Speaking Communities.....	16
Figure 3 Ethnolinguistic Map of Peoples of Iran.....	23
Figure 4: Discourse Continuum	37
Figure 5 Specificity, Definiteness, and Genericity. (Adapted from Karimi (1990)).....	59
Figure 6 The Givenness Hierarchy (Gundel, Hedberg, Zacharski 1993).....	64
Figure 7 Level of implicational restrictiveness of the Givenness Hierarchy.....	64
Figure 8 Higher statuses are associated with most restrictive statuses, and lower statuses with the least restrictive	64
Figure 9 The Cognitive Statuses are Implicationally Related.....	65
Figure 10 Traditional Concepts Map into GH and are Extended	67

Abbreviations

ATTR	attributive
CL	number classifier
COMP	comparative
DEF	definite
DEM	demonstrative
DIM	diminutive
DIST	distal (pronoun)
DUR	durative
IMP	imperative prefix
INFNT	infinitive
LNKPT	linking particle
NEG	negative prefix
NOM	nominalizer
PC	pronominal clitic
PL	plural suffix
PREP	preposition
PRES	present
PRO	pronoun
PSPT	past participle
REL	relativiser
SBJN	subjunctive prefix
SG	singular suffix
SUPRL	superlative
1	first person
2	second person
3	third person
q(G)	voiced dorso-uvular consonant

1 Introduction

This thesis investigates the referential status of the Persian morpheme *ra*¹. I will consider the cognitive (i.e., memory and attention) status of referents of *ra*-marked expressions in discourse. I will consider the cognitive statuses signaled by nominal expressions of *ra*-marked expressions in discourse, thus contributing to further understanding the exact nature of this morpheme.

The motivation for this study is the fact that the academic literature on functions of *ra* is inconsistent in claims about the referential status of this morpheme as a definiteness or a specificity marker. The morpheme *ra* in Persian has attracted considerable attention from traditional grammarians and cotemporary linguists alike, both in Iran and outside of Iran. A DP in Persian can occur with or without *ra* (with phonological variants *ro* and *o*), as illustrated in (1) and (2).

- (1) A: čī be-het dad?
 what to-you gave?
 'what did s/he give you?'

 B: ketab ∅ be-hem dad
 book ∅ to-me gave-3sg
 's/he gave a book (some books) to me'

 note that "B" here cannot be interpreted as
 * 's/he gave **the** book(s) to me'

(2) A: čī be-het dad?
 what to-you give-PAST-3sg
 'what did s/he give you?'

¹ *ra* is a word whose sole role is grammatical. However, the exact role of *ra* remains a point of contention in traditional grammar and in linguistic literature.

B: ketab **o** be-hem dad
 book **ra** to-me gave-3sg
 'S/he gave the book to me'

Note that "B" cannot be interpreted as
 * 's/he gave a (some) book(s) to me

At first glance, based on (1) and (2), one may be tempted to conclude, as has often been proposed, that **ra** marks definiteness. It is also important to keep in mind that not every DP has the option of co-occurring with **ra**. There are indeed cases in which the presence of **ra** results in ungrammaticality. The examples in (3) and (4) below show that there are DP's without **ra** that can be situationally interpreted as either definite or indefinite.

In (3), A has just finished college and has plans to upgrade her living arrangements. A and B have not talked for some time. However, B knows that A is still living in a single room in the boarding house where she has been renting for the past 6 years, and that she is looking to buy a house, although she may have to consider a condo as a second option, if she cannot find a house at a price she can afford:

(3) A: æsbab keši næ-kærd-i?
 Furniture pull NEG-did-3SG
 'did you move [yet]?'

B: næ, hænuz dar-æm dombale xune mi-gærd-æm
 no still have-1SG after house DUR-look-1SG
 'No, I am still looking for a house.' (i.e., 'I am house-hunting')

B': * hænuz dar-æm dombale xune **ra** mi-gærd-æm

Note that the insertion of **ra** above as in B' (* *hænuz dar-æm dombale xune ra mi-gærd-æm*) makes the sentence syntactically (as opposed to situationally) ungrammatical (i.e., The sentence is ungrammatical in any context).

In (4), A is joining B at a party, and is having difficulty locating the house where the party is. As she is driving up and down the street where she thinks the house is, A calls B on her cell phone for the 3rd time to get directions:

- (4) A1: alo?
 'hello?'
- B: Mæhnaz? næ-resid-i?
 Mahnaz NEG-arrived-2SG
- 'Mahnaz [is that you]²? Aren't you here [yet]?'
- A2: næ, (se bar dowre xyabun gašt-æm, væli)
 no, (three times around street search-1sg, but)
- hænuz dar-æm dombale xune mi-gærd-æm
 still have-1SG after house DUR-look-1SG
- 'No, (I have circled around the [same] street three times, but) I am still looking for **the** house.'
- A2': * hænuz dar-æm dombale xune **ra** mi-gærd-æm

The examples above show that i) not all DP's can get **ra**, and ii) not all DPs must be marked with **ra** to get a definite interpretation.³

While there have been a number of overlapping claims about the function of **ra**, researchers agree: that **ra** can never occur with subjects, that it does optionally occur with direct objects, and that it can mark the referential status of a DP. Further, **ra** has been claimed to mark definite and/or specific direct objects (Phillott (1919), Lambton (1953),

² Brackets here indicate implied meaning not reflected in actual code.

³ What may be prohibiting **ra** from occurring here is the compound verb "dombāl gashtan." The idea that certain verbs may prohibit the occurrence of **ra** has been mentioned in passing in literature. More specifically, Mahootian (1997) states, without any further elaboration, that "the compound verb *dombal-e gashtan*, 'to look for' does not allow any object to take –ra, no matter how definite". While I agree with Mahootian, as I presume any native speaker judgment would corroborate this claim, the idea that certain verbs may not allow the occurrence of **ra** altogether needs to be further investigated.

Browne (1970), among others); presuppositions (Ghameshi 1996, 1997); adverbial phrases of temporal and locational expression (Windfuhr 1979, Karimi 1989, 1990); topic/focus (Peterson (1974), Dabir-Moghaddam (1992)); and, case-marking obliqueness⁴ (Karimi 1989, 1990). Thus, it has been pointed out that *ra* does not occur only with direct objects. Karimi (1990) notes the following three cases: (i) "... *ra* co-occurs with adverbial NPs modifying intransitive verbs", (ii) "... *ra* may appear twice in the same clause", and (iii) "... *ra* co-occurs with noun phrase arguments (as opposed to adverbials...) which clearly are not direct objects of the verb". She offers the following in support:

- (5)
- | | | | |
|-------------------------------|---------|-----------------|---------------------|
| mæn-o | be-h-em | mi - xænd-e | |
| me -ra | at-me | Pres-laugh-s/he | |
| 'As for me, she laughs at me' | | | [Karimi 1990, (16)] |

Karimi points out that the verb *xændidæn* ('to laugh') does not take direct objects and offers the following to support this⁵ claim:

- (6)
- | | | |
|----|------------------------|---------------------|
| a. | *mæn-o mi-xænd-e | |
| | me-ra Pres-laugh -s/he | |
| b. | *man mi-xænd-e | [Karimi 1990, (18)] |

While I will necessarily consider direct objects and verbal complements as background information in understanding basic morphosyntactic aspects of *ra*, the focus of this dissertation is on the referential status of this morpheme.

⁴ Karimi defines oblique as anything except the subject.

⁵ Note that the verb *xændidæn* ('to laugh') does have a causative counterpart, *xændandæn* ('to cause to laugh') which does take *ra*.

Researchers all agree that *ra* never occurs with a nominative noun phrase and that it can mark referential status. However, there is a discussion surrounding the referential status of this morpheme as a *definiteness* marker versus a *specificity* marker. The discussion is further complicated by the fact that *ra* can co-occur with *yek* (“one/a”), as well as with *i*, a morpheme whose status is debated as well; there is also disagreement over whether the suffix *i* marks *definites* or *indefinites*. In this dissertation I will be considering the cognitive statuses of DP’s that occur only with *ra*.

Some researchers claim *ra* (*ro/o*, in speech) marks direct objects for definiteness. For example, Lambton (1953) provides the following example in support of this claim:

- (7) ketab *ra* be mæn dad
 book **ra** to 1-sg gave-3sg
 ‘s/he gave the book to me’

Ghameshi (1997) also claims that a direct object can optionally take *ra*, but when it does, it signals definiteness. Compare (8) and (9) below:

- (8) ketab xærid-æm
 book bought-1sgS
 ‘I bought books.’⁶ [Ghameshi 1997, (1)(a)]
- (9) ketab **o** xærid-æm
 book **ra** bought-1sgS
 ‘I bought the book.’ [Ghameshi 1997, (1)(b)]

Ghameshi notes that *ra* can never occur with subjects. Compare the two examples above with (10) and (11) below:

⁶ (8) can also be interpreted as ‘I bought a book’

- (10) ketab oftad- \emptyset
book fell+3sgS
'The book fell'⁷ [Ghimeshi 1997, (2)]
- (11) * ketab **o** oftad- \emptyset
book **ra** fell+3sgS

While this does not prove that *ra* never occurs with subjects, the examples above do show that *ra* distinguishes a definite/specific from an indefinite/non-specific interpretation for objects, and that (at least some) definite/specific subjects cannot take *ra*. Moreover, to my knowledge, other researchers have not pointed to any cases in which subjects would take *ra*, and my own native intuition supports the claim that *ra* does not occur with subjects. Further, while the examples above illustrate that there are cases in which definiteness and direct object position are sufficient conditions for *ra*, and that definite subjects cannot co-occur with *ra*, the data studied in this dissertation shows that DPs in adjunct position can co-occur with *ra* as well.⁸

Nominal expressions marked with *ra* can co-occur with one or more other morphemes, each of which interact not only with specificity and definiteness, but also co-occur with:

i) the plural marker, whose primary function is to select a group out of genericity; ii) *-i*, a suffix whose status is unresolved as it has variously been claimed to mark definites and/or indefinites; and, iii) *yek* (“one/a”).

⁷ (10) can be interpreted as

but

- a) 'A book fell',
- b) # 'books fell' is strained

⁸ While researchers have talked about this issue using notions such as “direct object” or “indirect object”, it should be noted that such notions may not be clearly identified cross-linguistically. In this dissertation, I will use the notions of “complement” and “adjunct” instead.

A plural marker in Persian functions to select a group out of genericity. An unmarked noun in Persian can be interpreted as singular or plural, as can be seen below.

- (12) *æsb* *xærid*
 horse buy-PAST-3SG
 ‘s/he bought a (some) horse(s)’

However, the plural marker can only indicate “(some of) the horses”:

- (13) *æsb-ha* *amæd-ænd*
 horse-PL come-PAST-3PL
 ‘The horses came’

The nominal expression *æsb* (‘horse’) can co-occur with *ra* and *ha*:

- (14) *æsb-ha ra* *xærid*
 horse-PL *ra* buy-PAST-3SG
 ‘s/he bought the horses’

A nominal expression can also appear with *-i*, in addition to *-ha* and *ra*.⁹

- (15) *æsb-i* *xærid*
 horse-IND buy-PAST-3SG
 ‘s/he bought a horse’
- (16) *æsb-ha* *xærid*
 horse-PL buy-PAST-3SG
 ‘s/he bought some horses’
- (17) *æsb-ha-i* *xærid*
 horse-PL-IND buy-PAST-3SG
 ‘s/he bought some horses’
- (18) *æsb-ha-i ra* *xærid*
 horse-PL-IND *ra* buy-PAST-3SG
 ‘s/he bought some horses’

Furthermore, the word *yek/ye* (‘one’) can co-occur with *-i* and *ra*.

- (19) *ye* *æsb* *xærid*
 one/a horse buy-PAST-3SG

⁹ While some of the translations given here are the same, only further context would determine the exact nature of nuanced meanings in each example. For example, *æsb ha-i xærid*, in context could mean “what horses he bought!”

‘s/he bought one/a horse’

- (20) **ye** **æsb-i** xærid (ke sævareš beše)
 one/a horse-IND buy-PAST-3SG (that ride-3SG SUBJ-become-3SG)
 ‘s/he bought one/a horse (so that s/he may ride it)’

- (21) **ye** **æsb-i** **o** xærid (ke sævareš beše)
 one/a horse-IND ra buy-PAST-3SG (that ride-3SG SUBJ-become-3SG)
 ‘s/he bought one/a horse (so that s/he may ride it)’

Moreover, *ye(k)*, *-i*, *-ha* and *ra*, and a demonstrative as in (22) below. Also, note the use of *ra* twice, once in the main clause and once in the subordinate clause.

- (22) **Yek-i** æz (un) **æsb-ha-i** **ro**
 one from (that/those) horse-PL-i ra

 ke xærideh bud **o** foruxt
 that buy-PAST-PERFECT-3SG ra sell-PAST-3SG

 ‘s/he sold one of the horses s/he had bought’

Recall that conflicting accounts of both *-i* and *ra* have been observed and reported in literature with regards to specificity and definiteness. For example, while *-i* in (23) would usually be interpreted as indefinite, the one in (24) would be considered definite.

- (23) **æsb-i** xærid-æm ke mi-læng-id
 horse-**i** buy-PAST-3SG that DUR-limp-3SG
 ‘I bought a horse that limped’

 æsb-i ke xærid-æm mi-læng-id
 horse-**i** that buy-PAST-3SG DUR-limp-3SG
 ‘The horse that I bought limped’

Examples such as those presented above have led some researchers to claim that *ra* is a marker of definiteness, while others have claimed that it marks specificity.

The apparent contradiction in categorizing elements such as *ra*, and *-i*, as well as their co-occurrence with the plural marker and demonstratives prove problematic for frameworks such as definiteness, and specificity which are dichotomous and provide us with only mutually exclusive categories. In other words, it is possible that the problem

may have to do with the frameworks themselves. As we will see in the course of this dissertation, the Givenness Hierarchy framework adopted here will allow for more fine-grained categories that are implicationally related, resulting in the avoidance of such apparent contradictions. While the current work is concerned with DP's that are marked only with *ra*, data collected included 14 DPs that included $-i+ra$. While the study of these DP's is outside the scope of this dissertation, they are nonetheless discussed in Section 4.9.2 for two reasons: i) to demonstrate that DPs containing both morphemes clearly function differently from those that are marked only with *ra*, and, ii) to outline a course of study for future work beyond this current work.

This dissertation is organized as follows: In the remainder of this Chapter I will provide a brief background on the history of the Persian language, and consider some of the difficulties inherent in studying Persian grammar. Chapter 2 is dedicated to providing a sketch of Persian grammar in general, and further provides aspects of Persian morphosyntax as they pertain to the *ra* morpheme in particular. This brief sketch is to explicate some basic aspects of Persian grammar for the uninitiated readers. In Chapter 3 I will outline the methodology in data collection and analysis. In Chapter 4, I will consider the Cognitive Status of *ra*-marked nominal expressions in discourse, presenting the results of the study. Further conclusions are drawn in Chapter 5.

1.1 Background

This section provides a backdrop on the historical nature of the language studied here, and some preliminary background on the particular morpheme in question. Section 1.1.1 will give some brief background information on the history of Persian as an Indo-European language, and the influence of Arabic on its development. Considering that there are many varieties of Persians in use today, studying aspects of Persian grammar, a researcher must also establish which variety or varieties are under investigation. The existence and the evolution of a number of different varieties in close proximity over a long period of time may potentially present some difficulties in identifying “a” Persian to investigate. To this end, Section 1.1.2 will consider the ethno-linguistic diversity in Iran, and the role this diversity has had in the development of many spoken varieties throughout history. Once it is established that there are many varieties of Persian, I will briefly address the sources of my data for this dissertation¹⁰.

¹⁰ More detailed explanations will be given in the Methodology chapter.

1.1.1 History of Persian Language

“Persian” in its current use is not a monolithic language. There are many varieties of Persian in use today, not only on a dialectal (i.e., group) level, but also on an individual (idiolect¹¹) level. An understanding of the history of Persian and its development will help develop a better understanding of Persian in current use, and the difficulties it presents in studying Persian grammar.

The genetic relationship of Persian to Indo-European languages, despite surface similarities to Arabic, has been well-documented. The diagram below, based on Mahootian (1997) and the American Heritage Dictionary (3rd Ed.) shows the relationship between Persian and some of the major branches of Indo-European languages.

¹¹ The term ‘idiolect’ is used here in a restricted sense to mean “...the speech habits of a person as displayed in a particular variety at a given time.” (Crystal, 1997)

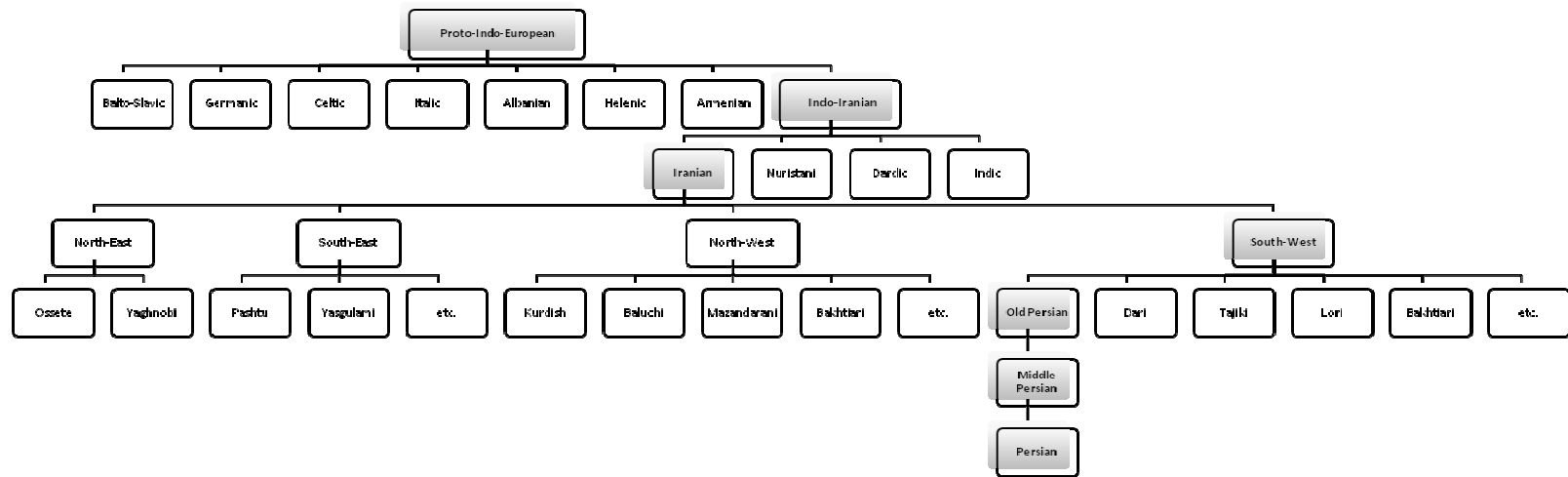


Figure 1 Some of the Major Proto-Indo-European (Circa 5000 B.C.) Languages.

As can be seen in Figure 1 above, Modern Persian is a descendent of Middle and Old Persian, a branch of South-West Iranian. Historically, early Iranian languages date back to 2000 B.C. The following timeline of Iranian languages in general, interacting with Persian (as an Iranian language) in particular is adapted from Windfuhr (2010).

Young Avestan 1000-600 BC	Early Middle Iranian 600-300 BC	Middle Iranian 300 BC-700 AD	New Iranian 700 AD
Old Persian		Middle Persian	New Persian

Table 1: Timeline of Iranian Languages Intersecting with Persian

Iranian languages¹² have “an estimated 150 to 200 million native speaker [constituting] one of the world’s major language families” (Windfuhr 2010:1). These languages are spoken today in linguistic enclaves in a large geographical area, extending from Syria to Turkey, Azerbaijan, Iraq, Iran, Pakistan, Afghanistan, Uzbekistan, Tajikistan, and the western edge of China.

The modern varieties of Persian in today’s Iran are traced to Middle Persian (3rd B.C. to 7th centuries A.D.) and Old Persian (6th and 5th centuries B.C.). Old Persian was written in Cuneiform, several dialects of which are preserved on clay tablets. The oldest existing document of Iranian Languages is the Old Persian inscriptions on a rock face near the

¹² The phrase “Iranian languages” here refers to a language group, in the genealogical sense, as a member of Indo-Iranian languages, as a sub-group within the Proto-Indo-European languages. As such, it may be compared to, for example, “Romance languages”. Alternatively, I use “languages of Iran” to refer to languages spoken within the national boundaries of today’s Iran, formerly known in the West as Persia, to include Indo-European languages such as Persian, Kurdish, Baluchi, Mazandarai, or Turkic languages such as Azari, etc. Note that Iran’s (Persia) national boundaries once extended to North Africa in the West, to India on the East and southern parts of the former Soviet Union.

city of Kermanshah in the Zagros Mountains in Eastern part of today's Iran (Lazard 1975; Windfuhr 1979, 2010; among others). These inscriptions date back to 522 BC.

Middle Persian, also known as Pahlavi¹³, was written in the Aramaic alphabet. There are vast phonetic and grammatical differences between Old Persian and Middle Persian (Lazard 1975, Windfuhr 1979, among others). Modern Persian, known as "Farsi" ("Parsi")¹⁴ to Iranians, is primarily spoken in today's Iran as a native language to

¹³ The word *Pahlavi*, literally, meaning "of power," is derived from Old Persian *Pārti/Pārtian* ('Parthian', as in the "Parthian Dynasty" (247 BC - 224 AD)). According to the Dehkhodā Dictionary the term is derived from *parsov(w)eh* < *Parhoveh* < *palhoveh* < *pahleh*. The term was adopted by Reza Khan (Reza Shah (king) Pahlavi), who founded the Pahlavi Dynasty (1925-1979) in Iran.

¹⁴ It is important to point out that both 'Persian' and 'Farsi' refer to the same language. 'Persian' is the English word for this language, and 'Farsi' is the Persian language word. There is some insistence on the part of a vocal minority of Iranian expatriates for English speakers to use 'Farsi' instead of 'Persian'. This insistence has created confusion among English speakers, as well as within the bilingual Persian-English community. An analogy might be German expatriates insisting that English speakers or bilingual German speakers replace the word 'German' with 'Deutsch' while speaking/writing in the English language. Also, note that Farhangestan, the organizational language policy and planning body in Iran (an organization similar to *L'Académie française*), at the request of foreign governments for clarification, has frequently stated the following fact: 'Persian' is an English word, and 'Farsi' is a Persian word, both referring to one language. Also, note that the language was originally spoken by a group known as *Pārsiān* ("those of the Pars"). *Pārsiān* (today known as "Persians") were an ethnic group from a province in today's southern part of Iran known then as *Pārs* (today's 'Fars' Province), and their language was known as '*Pārsi*' ("of *Pārs*"). According to Oxford English Dictionary (OED), the original Persian word '*Pārsiān*' came to English via Hellenistic Greek (<Middle English < Middle French < Classical Latin < Hellenistic Greek < Old Persian *Pārsian*) and has remained largely intact since. However, these words were mispronounced as '*Fārs*', '*Fārsi*' and '*Fārsiān*' (and, assimilated into Arabic syllable structure in the case of *Faris*) by the conquering and ruling Arabic speakers who did not have the bilabial plosive in their native Arabic. This change, among many others that left Persian a substantially altered language, remained after approximately 300 years of Arab rule and domination in Iran over a thousand years ago. The linguistic situation during this period of time, and the extent to which it left Persian altered, is somewhat different from most other language contact situations in that use of Arabic completely replaced Persian in most, if not all domains of language use, except in the 'home' domain, for nearly 300 years.

approximately 50% of the population. Other varieties of Persian are spoken as a native language by approximately 30% of the population in Afghanistan and 30% of the population in Tajikistan. Furthermore, Persian language enclaves can be found along the Silk Route from the Mediterranean Sea to India and China, as can be seen in the following map, taken from Commons Wikimedia
(http://commons.wikimedia.org/wiki/File:Iranian_languages_area.png (3/11/2010)).

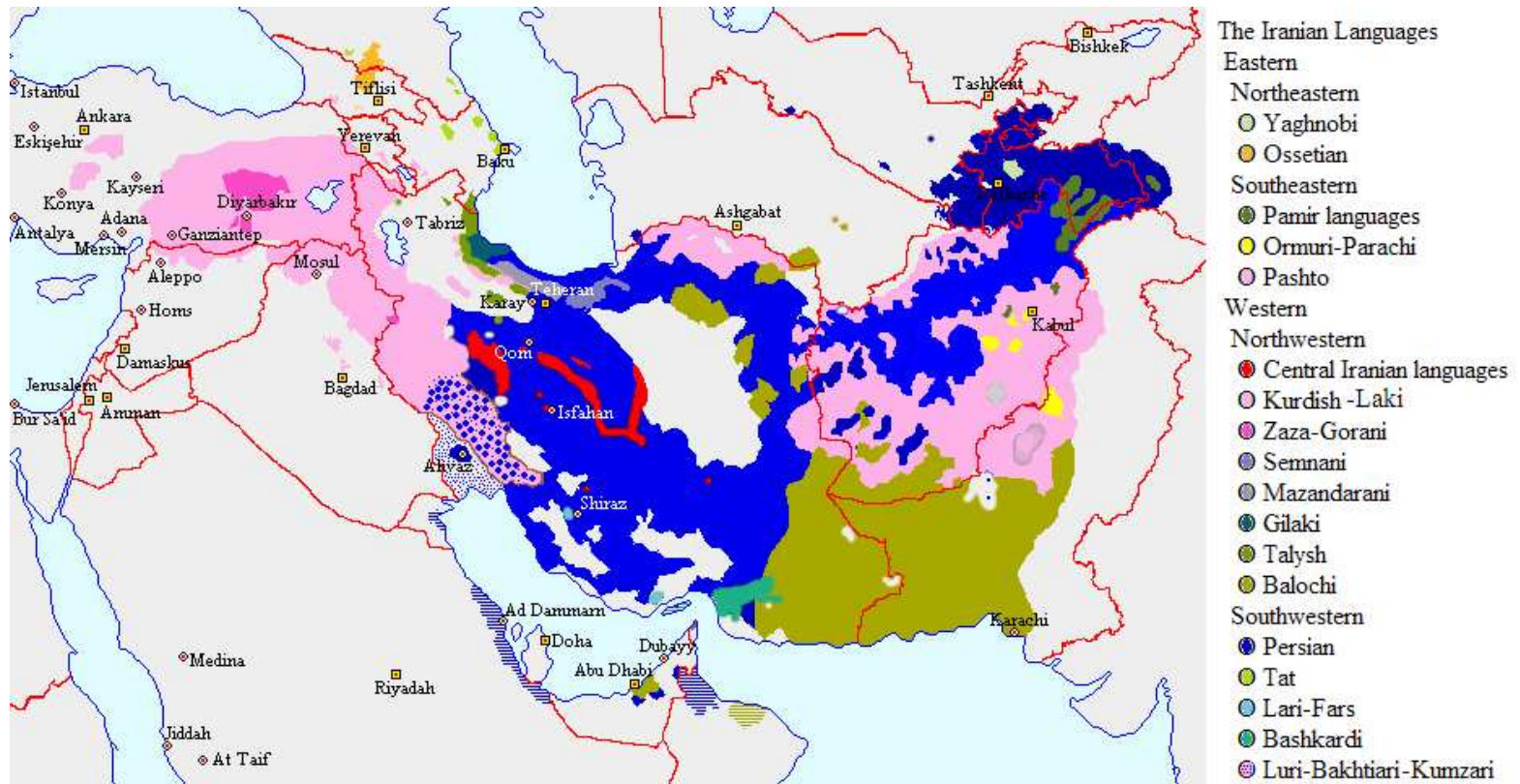


Figure 2 Map of Iranian Speaking Communities

As can be seen on the map in Figure 2, while the majority of Persian speakers can be found primarily in Iran, Afghanistan, and Tajikistan, there are also large enclaves in Pakistan, Turkey, Iraq, and the Caucasus. While this map shows Persian and some of the other major Iranian languages in the region¹⁵, there are many non-Iranian languages spoken in the region as well, interacting and influencing each other in language contact situations throughout history. For example, the influence of Azerbaijani (a Turkic/Altaic language) which has the most number of speakers next to Persian in Iran is noteworthy (Mahootian, 1997:1). However, perhaps the single non-Iranian language that has influenced Persian the most is a Semitic language: Arabic.

Historically, Modern Persian overlaps with the Arab conquest and the Islamization of the Persian speakers in the seventh century A.D.¹⁶ While written records are all but non-existent from 7th to 9th centuries (during the Arab conquest), what is regarded as Modern Persian in Iran begins to appear in Arabic script sometime between the 9th and 10th centuries¹⁷. The Arab conquest of Iran left very little behind from pre-Islamic times. The

¹⁵ While Figure 1 and Figure 2 show some of the Iranian languages spoken today, there are a number of other major Iranian languages in the region. For example, Mahootian (1997) lists the following: Persian, Dari, Tajiki, Luri, Bakhtiari, Kumzari, Kurdish, Talishi, Baluchi, Gilaki, Mazandarani, Zaza, Gurani, Bashkardi, Ormuri, Semnani, Zanjani, Saveh, Vafsi, Ashtiyani, Pashto, Yazgulami, Roshani, bartangi, Oroshori, Sarikoli, Sanlechi, Zebaki, Wakhi, Mujni, Yidgha, Ossete, Yaghnobi, as well as several dialects of central Iran.

¹⁶ The extent to which this overlap has influenced the local varieties depends on sociolinguistics factors, including historical, cultural, political, and religious influences. The subject falls outside the scope of this dissertation. For further discussions see Lazard (1975)

¹⁷ For further information on the scarcity of original linguistic materials, and the difficulties in pinpointing different stages of the development of Modern Persian, including the period after the Arab conquest all the way up to “classical period” (13th – 15th centuries A.D.) Readers are encouraged to consult Windfuhr (1979; 2010) and Lazard (1975).

great institutions of the country, including its monarchy and the Zoroastrian religion “were swept away by the Islamic conquest, and within three centuries there was little apparent remnant of them” (Axworthy, 2008). While most, if not all the institutions of Iran were destroyed, Persian as a language survived; however, the varieties of Persian that survived were infused with Arabic, much like English was changed by the influence of French after the Norman Conquest in the Eleventh Century. Furthermore, the variety that eventually became the standard literary language was the Dari dialect¹⁸.

Following the Arab conquest of Iran, which included cultural, religious, and political domination, the Persian language underwent considerable changes, as it was sent underground for approximately 300 years (Lazard, 1975). While there were a number of distinct varieties of Persian that existed by the time of the Arab conquest, one variety, known as Dari, gained prestige and dominance. Linguistically speaking, while the domination of Arabic meant changes in the lexicon, other language subsystems were impacted as well (Windfuhr 1979; Lambton 1953; among others). Note that this situation is similar to others in similar extreme cases of language contact.

In addition to many of the lexical items and some morpho-syntactic features borrowed from Arabic, modern Persian has lost much of its inflections on nouns, adjectives, and verbs that were present in Middle and Old Persian. Literature, poetry and grammatical structures were influenced by Arabic, as the language continued to shift and change.

¹⁸ Note that Dari, in its spoken form, is used today primarily in Afghanistan. As such, the standard and literary Persian in use today in Iran is closer to the spoken dialect in Afghanistan than to that of any of the dialects spoken in Iran, not withstanding those spoken in the eastern parts of Iran where they are geographically closer to Afghanistan.

Windfuhr (1979:80) states that “a particularly interesting, if complex, problem is the Arabic loan component in Persian. It consists not simply of lexical entries... but... of a considerable number of morphological, and some syntactic, rules of Arabic. This holds not only for literary Persian but also, to a certain degree, for colloquial”. However, while Arabic is a major influence on literary, formal and informal Persian, it is not the only language that has affected modern Persian.

This section, so far, has provided some basic information on the genealogical relationship of Persian to Indo-European languages, and has identified Arabic as a major influence as a contact language. Nevertheless, the Persian varieties used today are influenced not only by Arabic, but also by other languages as many dialects of Persian have been developing in this region over the centuries. Turkish, Russian, Mongolian, and other languages native to this region have influenced contemporary Persian in different regions of Iran to varying degrees. Moreover, there has been some influence from languages not native to this region as well. For example, French and English as International languages have been a source of borrowings over the past two centuries¹⁹. For example, the word ‘merci’ is used by many, if not most, Tehranis to say “thank you”. Alternatives are ‘motšækkeræm’ or ‘tæšækkor’ from Arabic, and perhaps less often, the Persian ‘mæmnun’. Similarly, the word ‘kop’ is used to mean “cup”, but only in the south-western part of Iran, near the oil fields where the British have had some influence during

¹⁹ Note that many Iranians sent their children abroad, beginning in early 1800’s to be educated in English and French speaking schools. As these educated students from prominent, influential families returned, they brought back with them a way of speaking that was to be emulated by others. More on education and language use will follow in section 1.1.2.3.

the 20th century. Since there are a variety of Persians in existence today, the situation may present some difficulties for the investigator.

In the next section I will consider the role of ethno-linguistic diversity in Iran in the development of various spoken varieties, and the role of education in spreading a particular dialect as the standard variety. These two factors have resulted in a diversity of Persian varieties observable in speakers across the entire Persian speaking communities within Iran. A grammar of Persian, then, necessarily must address the question of *which* variety it is attempting to describe.

1.1.2 Difficulties Investigating Persian Grammar: Would the “Real” Persian Please Stand Up?

Studying Persian grammar presents a number of difficulties for investigators²⁰. Among them is the diversity of the varieties called “Persian”, and the differences within a particular variety of Persian in use today. This is due to a number of reasons, including geographical factors; migration patterns of various Persian, as well as non-Persian speaking groups; and social factors to include the role of education and standardization in Iranian society. The issue is further complicated by the fact that extensive variations exist between spoken, written, and literary forms, as well as levels of formality within a particular context, where informal, formal, and literary features may be blurred in a single discourse situation. These complex linguistic interactions in Iran have resulted in extensive dialectal and idiolectal differences. When considering the different varieties in use in Iran, at its most basic level, we may distinguish between spoken varieties versus a standard written form. The ethno-linguistic diversity in Iran has resulted in the emergence of a considerable number of locally spoken varieties. These local varieties of Persian may exhibit differences, to varying degrees, not only from the standard variety, but also from each other.

²⁰ Many aspects of Persian grammar remain unresolved (Lazard 1975; Windfuhr 1979 & 2010; among others).

1.1.2.1 Ethno-linguistic Diversity

Traditionally, there have been many linguistic enclaves in the region that have existed in close proximity for centuries or millennia²¹. In some cases these enclaves have been separated by natural barriers such as rivers, subtropical forests, deserts, and mountains, and in other cases they have existed literally side-by-side. The following map taken from the library website at the University of Texas at Austin outlines the major ethnolinguistic areas in today's Iran and some of the surrounding regions²².

²¹ There are a number of different reasons for the establishment of these enclaves across the region. For example, throughout history, as news of invading armies from east or west would reach inhabitants of a region, entire peoples in towns and villages would either go underground, literally, or pick up and move, hitching rides on caravans along the many branches of the Silk Route. In some cases, they would settle in hard to reach areas such as valleys in the mountains where they would be away from potential attack. These settlements became permanent. To this day, there are settlements in the region that seem completely isolated from the world at large. There are many villages across the country that have no running water, no electricity, and the closest road may be hours away, requiring miles of hiking and pack animals to get to. Some of these communities have grown over time. My father was from a town called Golpayegan on the foothills of the Zagros mountains in western Iran. While it is not clear where the original inhabitants came from (or when they came), there are monuments in this town, dating back to the Saljuqi period in the 11th and 12th centuries, A.D. By the 1960's and 70's, there were approximately 22,000 people living in the town proper. Traveling from Tehran and back, a 400 mile trip, meant spending an hour and a half on a narrow dirt road to get to it. My father grew up in an old house, where there was a trap door on the floor of their stables, connecting the house to a labyrinthine "underground city". The story was that the structure was built to fool invading armies into thinking the town had been abandoned upon hearing the news of the coming soldiers, as the entire city moved underground.

²² It is important to recognize that while the Iranian national boundaries have been continuously shrinking for the past 2,500 years, into the 19th and 20th centuries, often the political redrawing of these borders have resulted in dividing and separating peoples of the same ethnic and linguistic background. One outcome of such divisions, as happens in similar situations around the world, has been the further bifurcations of established dialects.

Peoples of Iran

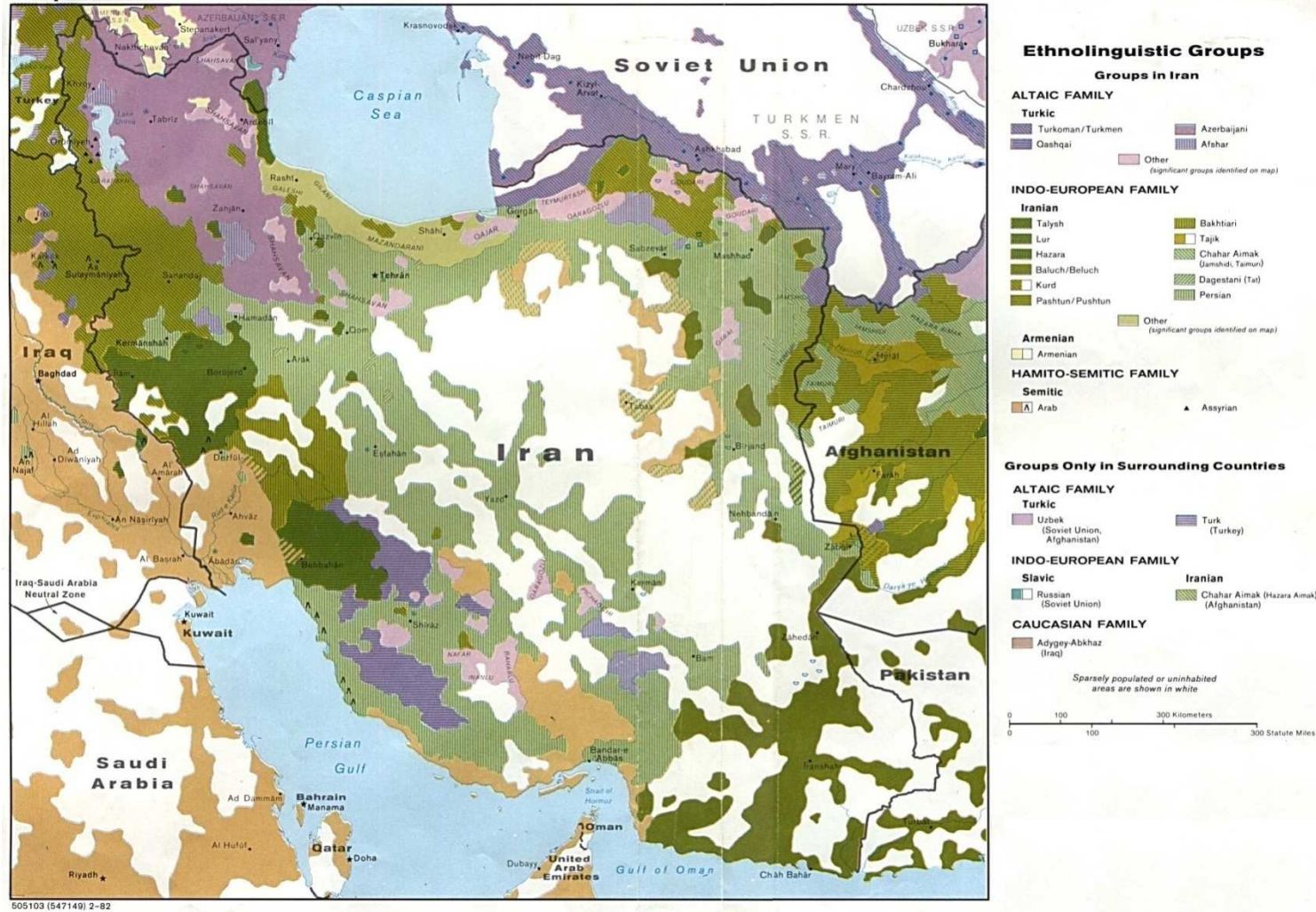


Figure 3 Ethnolinguistic Map of Peoples of Iran

Figure 3 shows a number of major language enclaves from several different language families existing in close proximity. The Iranian languages shown earlier on the map in Figure 2 (also, c.f., footnote 15) are still spoken in this region. In addition to the permanent settlements, migratory bands of tribes moved across the entire country, sometimes crossing national boundaries. These tribes spoke their own languages and dialects, by and large without written forms. As these migratory tribes moved from one place to another, their language continued to be enriched, as they in turn enriched the varieties they came into contact with. Throughout the centuries, however, governments of Iran have increasingly placed pressure on these migratory tribes to settle in one place. As a result, bands of tribes have largely settled around the edges of larger towns and cities, and in some cases, in small villages²³.

The settlement of a group of speakers in the midst of speakers of other language groups is not limited to the settlement of migratory tribes. Groups of peoples throughout history have had to pick up and move for various reasons, including war and invading armies. As a result, there is a long tradition of cities in Iran having linguistic enclaves. These enclaves, by and large, have continued to celebrate their own language and culture²⁴. In other words, while a city may have its primary language or dialect, it can have several

²³ For example, on a trip to Iran in 2006, I visited a small village just outside a city called Golpayegan on the eastern side of the Zagros mountains in eastern Iran. The village traditionally had been a Persian speaking community; however, large numbers of Luri speakers had moved into the village in the past 30 years, nearly quadrupling the population.

²⁴ There is a long-standing tradition in Iran of different languages and cultures coexisting. For example, in ancient times when the Persian armies conquered new regions, they left all institutions (religious, linguistic, etc.) intact. The newly annexed area was required only to pay taxes to the federal government.

major linguistic enclaves within its boundaries, much like, say, China Town in San Francisco. For example, while the city of Tabriz in northwestern Iran is primarily Turkish speaking, there are also Persian, Kurdish, Baluchi, and Armenian enclaves existing within the city. Conversely, one can observe Armenian or Turkish quarters in the middle of a larger Persian speaking community such as Tehran, Isfahan, or Shiraz. Each of these languages and dialects has contributed greatly and enriched the spoken varieties throughout each city, as well as the entire country.

1.1.2.2 Ethnolinguistic Diversity and Spoken Varieties

The coexistence of different peoples of Iran for centuries, and in some cases, for millennia, and the interaction between these people, has resulted in a number of different Persian dialects. These dialects may exhibit vast differences that have developed over the centuries. The spoken varieties used in linguistic enclaves, largely due to geographical barriers such as harsh mountainous regions and deserts, or social isolation, have continued to change, as spoken varieties do. Meanwhile, as mentioned earlier, the standard variety, codified in the works of the great Persian poets and writers after the 10th century A.D., has by and large stayed the same.

The chasm between a spoken variety and the standard variety is fundamentally no different than, say, English²⁵. At some level, this is no different from most other linguistic

²⁵ For example, in Standard American English (SAE), there is a two way distinction in the use of the copula, as can be seen in below.

- a. She is my friend
- b. She's my friend

While SAE exhibits a two-way distinction, in some varieties of English, a three way distinction in the use of the copula can be observed:

situations across all languages that have a written, literary tradition that has been codified into a standard form. However, what is different in the case of Persian is the fact that not only oral varieties can be vastly different from each other and from the standard variety, but also the number of different varieties that coexist in a relatively small geographical area.²⁶ While some of the differences from one speech community to another may be minor, there are communities with vast differences in their language of everyday conversation, each of which is incomprehensible to Persian speakers at large.

Furthermore, a speaker in a single discourse (or even a single phrase) may be codeswitching (e.g., Persian-Turkish, Persian-Armenian, etc.) and style-shifting, moving along a continuum, exhibiting features that span the entire spectrum from very informal, local speech to literary. Codeswitching and style-shifting, as exhibited in any bilingual speaking community, takes place while speaking to peers, with family, in public encounters, or interacting with positions of authority. Note that this speech by and large may be quite distinct from a more formal speaking style that is commonly used in broadcast news and public speaking. Note also that a literary style of speaking may be superimposed on any of these situations, to varying degrees, at any point in a single

-
- a. She is my friend
 - b. She's my friend
 - c. She my friend

²⁶ In this sense, perhaps a parallel situation might be Classical Arabic vs. Modern Standard Arabic vs. the many spoken varieties in use in Arabic speaking communities transcending national boundaries. Considering the presence of features from Middle Persian and Early Modern Persian (post Arab conquest) still in use in literary and formal speech and writing, the situation may be more comparable to that of Arabic and the influence of Classical (i.e., Quranic) Arabic in Modern Standard Arabic, juxtaposed with the different spoken varieties of Arabic. These varieties are used as primary language of every-day communication in communities across the globe, transcending national boundaries.

utterance. The language produced is complex, to extent that at times it is nearly impossible to separate the different styles within the same discourse: speakers tend to shift from one style to another within the same discourse, and at times within the same sentence.

The extent to which a particular speaker may exhibit shifting in speech, at some level, is related to how greatly their native variety differs from that of the prestige standard variety. For instance, those who may come from a socio-economic or educational background where their speech greatly differs from that of the educated, middle- to upper-classes, there is a target language they may be aspiring to produce, albeit not always entirely successfully. In the next section, I will present a brief history of the national educational system, the Iranian's attitude towards education, and role of modern Education in the development and the use of Persian in oral discourse.

1.1.2.3 Education and the Emergence of a Prestige Variety

The particular variety of Persian that gained prominence after the Arab conquest, as mentioned earlier, was Dari. Although substantial changes took place from the pre-Islamic to post-Islamic era, today's literary and formal standard²⁷ Persian is remarkably similar to that of the eleventh century Dari Persian. At the time of the Arab conquest, this dialect was used in the eastern part of Iran, which included today's Afghanistan. It was

²⁷ Note that while thus far I have been making references to a "standard" variety, in sociolinguistics it is understood that there is no one single form of a language called "a" standard. Even in the case of RP (Received Pronunciation), there are usually more than one single acceptable forms for language features. A "standard", then, by definition is a variety constituting more than one acceptable feature within certain restricted domains of language use.

this particular variety that came to be written down in poetry and literature, and eventually became the modern standard variety that has been used in Iran up to this day. As a result, the standard formal and literary variety used in schools across Iran today is closer to some of the spoken varieties in Afghanistan than to that of most of the regions in Iran.

Today, while not designated as the "official language," this particular variety of Persian is the national language of Iran and is used in many domains of language use, including government and education²⁸. Iran, however, is a multi-lingual, multi-ethnic country, as we saw in the previous section. Axworthy (2008) notes that "Iran is commonly thought of [by non-Iranians] as a homogeneous nation, with a strong national culture, but minorities like the Azeris, Kurds, Gilakis, Baluchis, and the Turkmen make up nearly half the population." While Persian is the native language of approximately 50% of school children entering the educational systems in Iran, as a lingua franca, it is spoken by virtually all Iranians, estimated at 70 million (2006, Iran Census).

Education and the educational system in Iran have had an enormous effect on the propagation of the literary dialect. Nevertheless, Iranians have always had an affinity

²⁸ Typically, in the field of Language Policy and Planning, certain distinctions are made between an "official" and a "national" language. Often, *official language* is used when a special legal status (typically associated with courts, parliaments, and administrations) is given to a particular language in a particular territory (country, state, or other jurisdictions). For example, in India, English is an 'official language' (in addition to Hindi), or Māori in New Zealand, in addition to English and New Zealand Sign Language. A *National Language*, however, is considered a language that has some connection with the people and the territory in which they live. In the case of Persian, while it is not officially declared the Official Language of Iran, as a 'national language' it functions as such. This situation is not dissimilar to that of the status of English in the United States.

towards learning and knowledge, which has included the memorization of lines of poetry written in the past millennia, even by the uneducated and the illiterate. The word *daneš* ('knowledge'), for example, shows up over and over again in poetry and in prayers. One would be hard pressed to find any Iranian today who would not be able to recite the following poems from memory.

- (25) ze gəhvare(h) ta gur daneš bejuy
 from cradle to grave knowledge IMP-seek-3sg
 'seek knowledge from the cradle to the grave'

Or, these lines from Ferdowsi's 1000 year-old book:

- (26) təvana bovəd hæɾ ke dana bovəd /
 capable is any that knowledgeable is

 ze daneš del-e pir borna bovəd
 from knowledge heart-of old young is

 'Capable is one who is knowledgeable/ knowledge makes old hearts young.'

The affinity towards knowledge and learning in Iranian culture goes much further back in history. Zoroastrians had a phrase known to all Iranians to this day: "good thoughts, good words, and good deeds" ('pendar-e nik, goftar-e nik, kerdar-e nik'). Menashri (1992) reports that the Achaemenids (550-330 B.C.) "stressed the value placed on knowledge", and that the Zoroastrians in their prayers asked Ahura Mazda, their supreme God, to grant them "educated children". Education has always been a high ideal for Iranians, both in pre-Islamic and in post-Islamic eras. Menashri also reports that the modern educational practices and traditions in today's Iran are based on this rich history of affinity towards learning, and poetry has played a great role in the propagation of ideas and ideals.

In 1963, the Shah ('King') of Iran introduced a set of reforms, called *The White Revolution*. These reforms, a set of political, social, and economic policies, contributed

significantly in the shaping of today's Iran. The aim was to bridge social gaps, and education played a major role to this end. Three of the 12 articles dealt specifically with "Educational Revolution" ('enqelab-e amuzeši'): The creation of a "Literacy Corps" (Article 6); education reform (Article 12); and, free schooling at all levels of education (Article 15). Menashri reports that "the most rapid expansion of the pre-university educational system occurred during the years of the White Revolution". In the 1960's and 1970's, an Education Corp (in addition to a Medicine Corps and an Agriculture Corps) were mobilized and deployed throughout the entire country. The results have been successful, with Iran today having one of the highest literacy rates in the world with women leading the numbers. This was a population that had developed their own spoken varieties in their isolated linguistic communities, sometimes with little or no contact with even their adjacent communities. While poetry as an oral tradition had kept the standard variety alive for the masses, as the population at large became increasingly educated, the use of a standard, literary form gained further prestige.

Menashri reports that "the prestige attached to higher education and the standing it gave to academics continued to rise". To this day, for example, forms of address such as Mr. Doctor ('aqa-ye doctor') and Mr. Engineer ('aqa-ye mohandes') are quite prevalent in marking honor and social standing. Note that these forms of address are used towards individuals not only in their profession, but also in their social circles and even within their immediate family. Interestingly enough, sometimes these forms of address are used even in situations where the individual addressed may not even hold such degree, but to pay them respect. The high regard with which education is associated in Iran is shared not only by the elite and the middle class, but also by individuals and families from all levels

of social strata in urban, rural or tribal settings. Menashri recounts that in his travels throughout Iran, he found that “education was held in high regard even in the most remote regions. The most striking example [he] came across was in the small village of Razun in Lorestan [a province in Western-central Iran]. The village had no running water, no electricity, no sanitation, not even a public ‘hammam’ (bath house). Yet when [he] asked the villagers what they needed most, they had only one answer: ‘The only thing we really need is a school. Only that can assure our children a better life than ours.’ The lower urban class had the same attitude.” (Menashri: 170). Note that the population in the Province of Lorestan predominately speaks Lori, a South-West dialect of Iranian languages (c.f., Figure 1).

The spread of Standard Persian, in part, is rooted in increased literacy, and the role education has in social mobility and concepts of prestige. Parents across all social strata throughout Iran seek to secure the best possible education they can for their children—the perception is that education is a main vehicle for penetrating social barriers to achieve status and power. Failure to achieve educational goals or to gain admission to certain schools, in some cases, has resulted in suicide, as education became the first and foremost vehicle for social mobility.

The language of the academy, of course, is singularly Standard Persian, regardless of the student’s ethno-linguistic background. As a predominately illiterate culture—with already existing high regard for poetry and the literary variety—Standard Persian gained an even higher standing in a relatively short period of time. The language of the academy in Iran—as is the case in other language communities with varieties different from the

codified²⁹ Standard variety—is considered somehow more elegant, superior, and associated with higher social standing. As the use of much-admired poetry and literary prose increased, the distinctions between local spoken varieties and Standard Persian continue to be blurred. Ordinary conversation at all levels of society has been influenced by academic discourse, as a direct, one-to-one relationship is assumed between higher social status and the use of this particular variety.

Historically, many spoken varieties across the land continued to develop and evolve among peoples with no written form of their own variety, and with no reading or writing skills. While the standard variety served in poetry recitation, and in formal and literary situations, spoken varieties diverged, partially due lack of contact among communities of speakers, and partially due to no reading and writing skills among the majority of speakers. Ethnic, social, and linguistic distinctions traditionally kept enclaves separated and largely intact over long periods of time. Today, there are many spoken contemporary varieties of Persian in use. Several major factors are interacting with each other at this point in time to influence the language under study here. First, these different varieties have traditionally been permeated with features from Iranian and non-Iranian languages. Second, the literary variety, remarkable unchanged, has been codified and kept alive, not only in writing by the few who could read and write, but also memorized in the form of poetry by those who could not. Finally, and perhaps more importantly, modern education has had a major impact on the spoken varieties. As a result, the language situation in today's Iran exhibits a degree of volatility. Again, fundamentally, this is no different than

²⁹ Codification here is used in the technical sense to mean a variety that has dictionaries, grammars, rules for “correct” usage, etc. Note that non-Standard varieties for the most part do not enjoy the same accord.

other language situations; the question is only that of the degree to which a language situation is stabilized.

Against this linguistic background, there are hosts of dialectal and idiolectal varieties in use today, ranging from colloquial varieties to standard formal and literary varieties that can not only be used in writing, but in spoken situations as well. Furthermore, a single speaker, in any given language context, spoken or written, may shift from one variety to another in a single discourse situation. A dialogue or a piece of writing may be layered with the full range of possibilities, blurring any distinctions between colloquial, formal, and literary features. This situation, of course, may present some difficulty in studying Persian grammar.

For the purposes of this dissertation, I have relied on data gathered from several different written varieties. In addition, I have chosen one particular spoken variety, Tehrani, to examine. In Section 1.2 below, I will briefly discuss the sources of data used in this study, and explain why Tehrani was chosen as the source for spoken data. Further detailed information on the actual data used is given later in the Methodology Section.

1.2 Modern Persian in the Context of the Present Study

The spoken style chosen for this study examines a variety of Persian that is spoken in Tehran. As a bi-dialectal speaker of Persian, I have native command of Golpayegani and near-native Tehrani³⁰. Additionally, and more importantly, Tehrani was chosen for this study because it is considered the most “neutral” (i.e., Unmarked), and the most prestigious spoken variety in Iran, as it is the most commonly heard speech on mass media (T.V., radio, movies, etc.).

Tehran grew from a small community to a metropolis in the course of about 200 years. In 1978, when I left Iran, it had a population of approximately 3 million people. Today, Tehran is home to approximately 10 million. Much of Tehran’s growth is due to migration patterns: many of the Tehranis have moved to this city from various parts of Iran to include large populations of Turks, Kords, Armenians, Afghanis, to name a few, and have enriched the Tehrani dialect with their contributions. It can be argued that Tehrani is the most widely used and most easily understood dialect to all Iranians, as it has relatively quickly become the most socially prestigious spoken variety used by the national media. Furthermore, this spoken dialect has become the modern standard written variety used increasingly in many written domains of language use, particularly in genres such as fiction and screen plays where dialogue is of importance. For these reasons, Tehrani is considered the most neutral dialect in relation to other Persian dialects.

³⁰ I lived and attended schools for the first 16 years of my life in Golpayegan and Tehran. I am also fluent in Esfahani, as one side of my family is from Esfahan. I grew up hearing the dialect, and spent extended

Note that while the spoken variety investigated here can be broadly referred to as the Tehrani variety (i.e., the variety most likely to be encountered in Tehran), this is not a singular variety as speakers in Tehran come from all over the Persian speaking communities. While in many cases spoken constructions can only be judged grammatical in extremely context-rich environments, at times it is difficult for native speakers, as well as researchers themselves, to come to a consensus.

Conclusion

In conclusion, due to the complex linguistic situation, the data collected for this study comes from a variety of sources that I argue best reflect this current state of affairs.

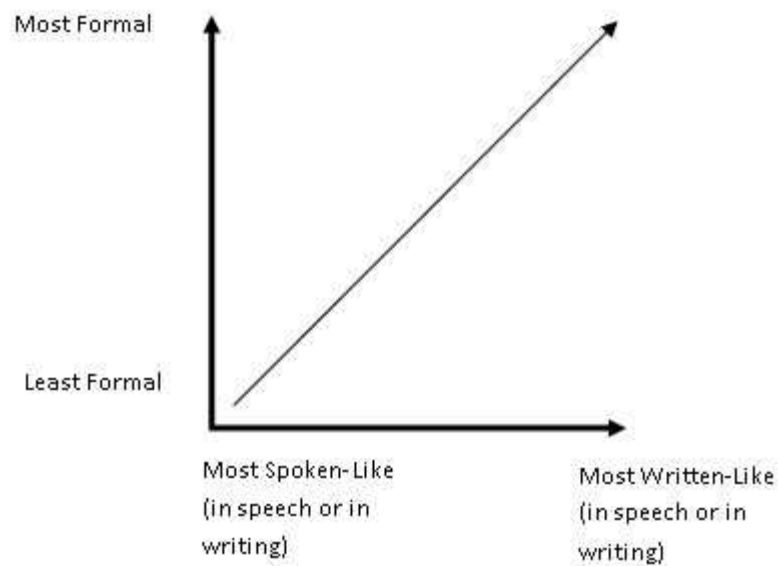
The linguistic analysis in this dissertation is, of course, impacted in several ways by the diverse usage as outlined in this section. The situation outlined here has further, overall implications as well. The situation impacts grammaticality judgments, the claims made in the literature about *ra*, and, of course, my own analysis. In other words, a given construction may be judged to be morpho-syntactically (un)grammatical, or pragmatically (in)appropriate, depending on which variety of Persian, which speaker, and which context we are dealing with.

periods of time in the city of Esfahan, including entire summers attending summer classes at the Association of Iran and America (*anjoman-e Iran va Amrika*)

2 Persian Grammar: A Sketch

In this section I will give a brief description of some basic concepts in Persian grammar.

A given phrase in Persian may have one or more syntactically possible and pragmatically appropriate options, including changes in word order, word choice, morphological markers, and grammatical categories (e.g., occurrence of present progressive only in speech, but not in formal written/literary phrases). As mentioned in previous sections, there may be many layers of usage, involving different levels of formality. At its most basic level, we may classify the language encountered (in speech or in writing) as spoken-like vs. written-like, or formal vs. informal. I offer the following diagram as a visual reference to the complexity that one encounters in Persian speech or writing.



Further binary distinctions, as well as distinctions on a continuum may be made in referring to a particular construction. Therefore, various terms and phrases such as *standard/non-standard*, *written/spoken*, *formal/informal* (*formal/informal written*,

formal/informal speech) *colloquial*, *careful colloquial*, etc. may be used to describe a particular utterance. While, potentially, we may place these distinctions on a continuum, it may prove difficult to draw discrete lines between each category and determine where a particular phrase (or constituent within a phrase) fits within such taxonomy.

Schematically, the situation may be represented as follows:

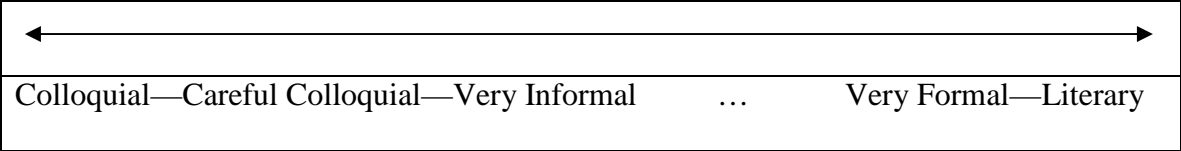


Figure 4: Discourse Continuum

The grammar described in this Section may be considered to describe the formal written language, unless otherwise stated. Constructed examples reflect the written standard language, and are transcribed to reflect this variety. Whenever dealing with speech, *careful colloquial* Tehrani is reflected in the transcriptions. Examples to describe other varieties of Persian are noted, as appropriate.

Persian is considered an SOV language, as can be seen in (27) below:

- (27) mæn danešju hæst-æm
 I student be-PRES-1SG
 'I am a student'

The prepositional phrase typically occurs after the subject and before the direct object (S-PP-O-V) as illustrated below:

- (28) mæn dær Minnesota danešju hæst-æm
 I in Minnesota student be-PRES-1SG
 'I am a student in Minnesota'

Canonical Persian word order is a mixture of head-final and head-initial strings: verbs occur finally, as in (27) and (28) above. Further, nouns precede adjectives (NA); prepositions precede nouns (PrepN); and nouns precede relativizers (NRel). Modifying elements are connected with a linking particle³¹ (“Ezafeh”). These constructions are exemplified below.

Noun and Adjective

- (29) æsb-e abi
 horse-LNKPT blue
 ‘blue horse’

Preposition and Noun

- (30) ru æsb
 on horse
 ‘on horse’

Noun and Relative Clause

- (31) sævar-i ke ru asb bud
 rider-DET REL on horse be-PAST-3SG
 ‘The rider that was on the/a horse’

Unbound demonstratives precede nouns, as demonstrated below:

- (32) an sævar
 DEM rider
 ‘that rider’

Plural markers follow the noun. However, plurals in Persian function to select a group from genericity. Thus, an unmodified noun in Persian can be interpreted as singular or plural, as in (33) below.

³¹ Head nouns and dependent nominals that follow are connected by *–e*. Traditionally this morpheme has been called the *ezafe* construction (borrowed from Arabic with similar, but not identical functions) by virtually all grammarians working with Persian, including modern linguists. The *ezafe* is multifunctional. In this dissertation I will use “linking particle (LNKPT)” to refer to this morpheme, as this is the function in question here. Thus, ‘linking particle’ here is used in the sense that a syntactic category is used to specify an attribute/link to the N it modifies, and in Persian grammar the linking/attributive particle is required to indicate this function.

- (33) æsb xarid
 horse buy-PAST-3SG
 's/he bought a (some) horse(s)' [i.e., 's/he engaged in horse-buying']

However, the plural marker indicates “some/the” horses:

- (34) æsb-ha amæd-ænd
 horse-PL come-PAST-3PL
 'The (some) horses came'

While the Persian word order is S-PP-O-V, in a given context, any or all the constituents, excluding the verb, may be omitted. That includes the subject, the object, and the prepositional phrase, leaving the sentence with only the verb: (S) (PP) (O) V.

- (35) danešju hast-æm
 student be-PRES-1SG
 'I am a student'

While the pronoun can be omitted, in formal, informal and colloquial situations, further contractions may take place. Compare (35) above to (36) below to observe the contraction of 'be' below:

- (36) danešju-æm
 student-1SG
 'I am a student'

Note that in speech, a further contraction of 'æ' may take place. A Tehrani, for example, may be heard to say³²:

³² Note that this type of contraction is not limited to Tehrani, or to this particular verb. Consider the following formal, standard Persian statement:

- 1) mæn ne-mi-dan-æm
 I NEG-DUR-know-1SG
 'I do not know'

In many of the dialects of Persian in Iran, one may observe the following change to above in speech:

- 2) mæn če mi-dan-æm
 I what DUR-know-1SG
 'I don't know' [i.e., 'What do I know!']

- (37) danešju-m
 student-1SG
 'I am a student'

verbs are not only marked for person and number, they are also marked for tense and aspect:

- (38) mæn be danešgah mi-ræv-æm
 I to university DUR-go-1SG
 'I go to university' ³³

Omissions and contractions may be involved in all levels of formality, ranging from literary to formal written, formal speech, careful colloquial, or more informal colloquial speech, as pragmatically appropriate. As expected, higher degrees of this phenomenon is exhibited in highly context-rich environment such as speech. So, (38), in Tehrani speech, may take on a variety of forms, as demonstrated below. In the following examples, note

In Golpayegani, a town in the Zagros mountains in the west-central part of Iran, this statement may be expressed as below:

- 3) mæn-xo čom-dun-æm
 I-REFL what-know-1SG
 'I don't know'

Similar changes can be observed in the following contractions in the Zarand dialect (in the Kerman Province, in south eastern part of Iran):

- 4) če mo-don-om
 what DUR-know-1SG
 'I don't know' ('What do I know!')

which can be further reduced:

- 5) če monom
 what I know

or, simply

- 6) chem

³³ Note that this utterance can be interpreted as "I go to university", "I am going to university", and "I will go to university". Also, note that 'university' is underspecified for definiteness/specificity (i.e., "a/the university"). Furthermore, 'university' is not marked for number (plural/singular) and as such may indicate one or more universities.

not only the omission of various constituents, but the reduction in the verb stem itself (syllable simplification: [ræv] → [r]). The *null* symbol (‘Ø’) is used here to show omissions.

- | | | | | |
|------|-----|----------------------|------------|------------|
| (39) | Ø | be | danešgaØ | mi-rØ-æm |
| | | to | university | DUR-go-1SG |
| | | ‘I go to university’ | | |
| | | | | |
| (40) | mæn | Ø | danešgaØ | mi-rØ-æm |
| | I | | university | DUR-go-1SG |
| | | ‘I go to university’ | | |
| | | | | |
| (41) | Ø | Ø | danešgaØ | mi-rØ-æm |
| | | | university | DUR-go-1SG |
| | | ‘I go to university’ | | |

Furthermore, while in formal constructions the verb is last, this may change as well.

Compare the following Tehrani utterance to the ones above.

- (42) (mæn) mi-r-æm (be) danešga
(I) DUR-go-1SG (to) university
'I go to university'

Note also that the direct object can be omitted as demonstrated in (43) below.

- (43) A: (emruz) ne-mi-r-i danešga?
(today) NEG-DUR-2SG university?
'Aren't you going to the University [to school] (today)?'
- B: (čera) mi-r-æm
(yes) DUR-go-1SG
'(yes) I'm going'

The following exchange from my data further illustrates some of the deletions that may take place in naturally occurring contexts. Note the object head drop ('Mehrak's *husband*') in (44) and (45), while we observe several levels of deletion in (45) B.

- (44) A: *rasti, šowhær-e čiz-o did-i,*
truth-DEF/IND husband-ASSOC thing-ra see-2SG

Ø *mehrak-o?*
Ø *Mehrak-ra*

‘By the way, did you see so-and-so’s husband, Mehrak’s [husband]?’³⁴
- B: Ø *mehrak-o? bæle ...*
Ø *Mehrak-ra yes*
‘Mehrak’s husband]? Yes.’
- (45) A: *četor bud?*
how was-3SG
‘How was [he]?’
- B: Ø *did-am[-Ø].* Ø *bæd ni Ø... ej,*
See-PAST-1SG Ø bad NEG [-is-3sg] INTERJ

bæd-æk ni Ø
Bad-DIM NEG[-is-3sg]

‘I saw [him]. [he] is not bad. [he] is not too bad.’

Outside the context, the object-head-drop could be misinterpreted as ‘Mehrak’ rather than her husband. In (45) B, several deletions are taking place: the object pronoun (“him”), the subject pronoun (“he”), as well as the reduction in the negated verb (*nist* → *ni*).

Recall that our concern here in this dissertation is nominal expressions containing *ra*.

While we saw that many of the elements can change form, they may be entirely omitted.

In the next section, I will consider *ra* in the context of the current work.

³⁴ Note that *čiz* (‘thing’) is used here as an interjection.

2.1 The Morpheme *ra*

In this section, I will consider some of the morphosyntactic, semantic, and pragmatic functions of *ra*. A DP in Persian can occur with or without *ra* (*ro/o*), as the following examples illustrate.

- (46) A: *či* *vasæt*³⁵ *xærid?*
 what for-you buy-PAST-3SG?
 ‘what did s/he buy for you?’
- B: *ketab* *ø* *vasæm* *xærid*
 book *ø* for-me buy-PAST-3SG
 ‘s/he bought a book (some books) for me’
- (47) A: *či* *vasæt* *xarid?*
 what for-you buy-PAST-3SG
 ‘what did s/he buy for you?’
- B: *ketab* *o* *vasæm* *xarid.*
 book *ra* for-me buy-PAST-3SG
 ‘S/he bought **the** book(s) for me’
 [e.g., ‘the book (that you know about) I had asked her to get for me’]

The morpheme *ra* in Persian has attracted considerable attention from traditional grammarians and contemporary linguists alike, both in Iran and outside of Iran. Windfuhr (2010) states that in constructions in Iranian languages, in the past, “the direct object/patient was in the nominative and later in the direct case, many languages disambiguated the role of the direct object by the strategy of grammaticalizing selected adpositions.” What is interesting to us, in the context of the present study, is that, “[e]ven neighboring and closely related dialects may differ widely in the progression of ...their direct object marking” (Windfuhr: 33). Regarding the morpheme *ra*, Windfuhr observes three stages of development: Old Persian, *radi*; Middle Persian, *ray*; and *ra* in Modern

³⁵ Note that the use of *vasæt* instead of the standard *bæra-ye to* ‘for you’ or the more formal/polite *bæra-ye šoma* ‘for you (plural)’ marks this utterance as specifically that of a Tehrani.

Persian. In the first stage, in Old Persian, *radi*, was used to mean ‘on account of’, ‘for the sake of’, or ‘concerning’. In Middle Persian, the form changes to *ray*, connoting a wide range of dative functions to include possessor with a copula. In addition to marking possession, *ray* also i) indicated beneficiary ‘for’, ii) marked indirect objects and as such, it alternated with the preposition ‘ō’ (“to”), and iii) was an “occasional marker of direct objects”. Windfuhr states that “...after the loss of the oblique case [some time during Old Persian into Middle Persian], the direct object remained regularly unmarked, and only in late texts was it occasionally marked by the postposition *rāy*, clearly under influence of Early New Persian” (Windfuhr 2010: 34). He adds that all of these functions in Middle Persian continued into Early New Persian, and that “[t]oday, *rā* has become the obligatory marker of the specific direct object, both definite and indefinite.” (Windfuhr 2010: 33-34).

The historical aspects of the postposition in question here are important in the context of this dissertation, to the extent that there are remnants of these ‘archaic’ functions still in use in today’s Persian, both in very formal and literary genres, as well as colloquial speech. In the socio-linguistic context of Persian in use today, many speakers of different varieties of Persian, like speakers of non-standard dialects of any language, tend to approximate the perceived “correct” target. For example, as mentioned earlier, one function of *ra* in Middle Persian (*ray*) was to mark possessor with a copula, as in (48) below.

- (48) *væzir* **ra** *doxtær-i* *bud*
 vizier *ra* daughter-IND be-PAST-3SG
 ‘there was a daughter to the vizier’ [i.e., “The vizier had a daughter”]

Today, while speakers of Persian would understand (48), and recognize it as a literary construction, (49) below would be its equivalent in more modern literary form³⁶:

- (49) vǝzir doxtǝr-i dǎšt
 vizier daughter-IND have-PAST-3SG
 “The vizier had a daughter.”

Similarly, remnants of the Middle Persian *ray* as the regular marker of indirect objects, alternating with the preposition ‘to’ can still be found in today’s Persian speakers’ linguistic knowledge, as in the Persian adage in current use in (50) below.

- (50) LoG³⁷man **ra** goftǝnd ǝdǝb³⁸ ǝz ke amuxti
 Loghman **ra** say-PAST-3PL manner from who learn-PAST-2SG
- goft ǝz bi-ǝdǝban, ke hǝr-čǝ ǝz
 say-PAST-2SG from without-manner-3PL that anything from
- ʔišan dǝr nǝzǝrǝm napǝsǝnd
 PRO-3-PL (SG polite) PREP viewpoint-1-POSS-SG NEG-admirable
- amǝd ǝz feʔl-e an
 come-PAST-3SG from act-LNKPT PRO-DIST
- pǝrhiz kǝrdǝm.
 abstinence do-PAST-1SG

‘(They) asked Loghman from whom did you learn [good] manners [i.e., how did you become cultured]? He responded, ‘from those without it, in that I abstained from anything I observed of them that I disapproved.’

³⁶ Note that a less literary, yet still formal version below.

vǝzier **yek** doxtǝr-i dǎšt
 vizier one daughter-IND have-PAST-3SG
 “The vizier had a daughter”

³⁷ Voiced dorso-uvular

³⁸ word ‘ǝdǝb’, a borrowing from Arabic, has a range of connotations in Persian. It is used to mean *manner, decorum, complaisance, curtsy, gentry, knowledge, culture*, or even *literature*, in the plural (i.e.,

Note that the phrase ‘Loghman **ra** goftænd’ in today’s Persian would be

- (51) be Loghman goftænd
 to Loghman say-PAST-3PL
 ‘they said to Loghman’

However, what is interesting is that if one walked up to an average person on the streets in most places in Iran and began the phrase “Loghman ra goftænd...,” the addressee would be very likely to know and finish the sentence “... ædæb æz ke amuxti? Goft æz bi-ædæban.” The point here is that while today this is a literary usage of this type of construction (and speakers may recognize it as such), it is still a part of today’s Persian speakers’ linguistic knowledge. Although outside the scope of this current work, it might be argued that this type of construction may be a part of at least some speakers’ attempt to approximate some “target,” and therefore impact one’s grammaticality judgment. In other words, there are “archaic” constructions that are still part of contemporary speakers’ mental grammar, and as such, a linguistic analysis must necessarily take these into account.

Iranian grammars (e.g., Razmjoo, Rokni et.al. 1992) have traditionally placed the morpheme *ra* in a category called “Neshaneh”, a term literally meaning ‘marker’. Research informed by modern linguistic approaches often refers to *ra* as a ‘postposition’ (Browne, 1970; Dabir-Moghaddam, 1992; Karimi, 1989; Windfuhr, 1990) or simply a ‘morpheme’ (Ghomeishi, 1997). While ‘morpheme’ is not mutually exclusive with ‘postposition,’ the term ‘postposition’ does suggest that it is not a bound morpheme. The examples below at the very least demonstrate that *ra* can be separated from the DP

‘ædæbiyat’, occurring with the Arabic plural marker, a somewhat common occurrence, particularly in Arabic borrowings)

‘apartment’, making it difficult to argue that it is a bound morpheme³⁹. Here I will simply refer to it as a morpheme and will remain neutral as to its status as a postposition.

- (52) apartemani **o** ke didim xæridim
 Apartment **ra** REL saw-1PL bought-1PL
 ‘We bought the apartment that we saw.’
- (53) apartemni ke didim **o** xaridim
 Apartment REL saw-1PL **ra** bought-1PL
 ‘we bought the apartment that we saw’

Aside from terminological issues, the functions attributed to *ra* are varied. Different analyses in current research on *ra* can be primarily divided into several overlapping claims: *ra* as a case marker, *ra* as a topic marker, and *ra* as a marker of reference (definiteness or specificity). As a morpho-syntactic category, traditionally *ra* has been assumed to mark accusative case. Below I will present a basic sketch of the claims about functions other than marker of reference that have been attributed to this morpheme. More specifically, I will consider concepts such as direct object, transitivity, and accusativity.

³⁹ Of course, this alone is not enough to support the argumentation as for example the English possessive morpheme ‘s’, e.g., “The man in the blue shirt’s uncle”.

2.1.1 Direct Objects and *ra*

Researchers agree that *ra* can mark direct objects, although there is some debate over whether they mark direct objects that are definite, specific, or definite and specific. To help readers who may not already be familiar with *ra* and its relationship to dependent elements of the verb structure, this subsection and the subsections that follow outline the connection between *ra* and direct objects.

The number of syntactic elements required by a verb directly affects the occurrence of *ra*. While subjects never get *ra*⁴⁰, direct objects optionally occur with this morpheme. Thus, verbs that function intransitively do not co-occur with *ra*. Examples of these types of verbs include *dævidæn* ('to run') and *amædæn* ('to come').

- (54) aræš amæd
 Arash come-PAST-3SG
 'Arash came.'
- * aræš **ra** amæd

In (54), 'Aræsh' is the subject and as such it does not get *ra*. Furthermore, there is no object to which *ra* can attach. In (55) below, the object in the prepositional phrase *be xane* ('to house') cannot be marked by *ra* because it is not a direct object (i.e., a direct complement of the verb).

- (55) aræš be xane amæd
 Aræsh to house come-PAST-3SG
 'Aræsh came to the/a house'
- But, not
 * aræš be xane **ra** amæd⁴¹

⁴⁰ That *ra* does not occur with subjects is universally agreed upon by researchers—perhaps one of the only facts about this morpheme that everyone agrees upon.

⁴¹ Note that in Persian a preposition can optionally drop in a PP, yet still no *ra* is allowed:

Conversely, transitive verbs require a direct object, which may be optionally marked with *ra*. Transitive verbs include verbs such as *bordæn* ('to take').

- (56) æsb *o* mi-bær-æm (tu tævile)
 horse *ra* DUR-take-1SG (in stable)
 'I am taking/ will take the horse (to the/a/some barn)'

With ditransitive verbs such as *dadæn* ('to give'), while both a direct object and an indirect object are required, only the direct object may take *ra*. Observe:

- (57) aræš æsb *o* be rostæm dad
 Arash horse *ra* to Rostam give-PAST-3SG
 'Arash gave the horse to Rostam.'

* aræš xane *ra* amæd

This is also true in the spoken version of this phrase in Tehrani, with different pronunciation and different word order (verb-second as opposed to verb-last):

aræš ʔumæd xune
 Arash came-PAST-3SG house
 'Arash came to the house'

but not:

*aræš ʔumæd xune *ro*

Note that if the verb *didæn* ('to see') is added to this phrase, it would be syntactically well-formed:

Aræš ʔumæd xunæ-ro be-bin-e
 Arash came house-ra SUBJ-see-3SG
 'Arash came to see the house.'

As illustrated in the examples above, *xane* ('house') cannot be marked with *ra*, with or without the preposition *be* ('to'). Note that this is regardless of its referential status. Observe below that *dævidæn* ('to run') works similarly to *amædæn* ('to come'):

æsb dæv-id
 horse run-PAST-3SG
 'the/a/some horse ran'
 * æsb *ra* dævid

- (58) aræš æsb be rostæm dad
 Arash horse to Rostam give-PAST-3SG
 ‘Arash gave a/some horse to Rostam.’

But not

*be Rostæm **ra**

As we have seen in this section, direct objects can optionally take *ra*, while subjects never do. This has been well established and well documented in the literature.

2.1.2 Transitivity, *ra*, and Ambiguity

Verb transitivity, at its most basic level, refers to a relationship that may hold between a verb and the dependent elements in its phrase structure. Intransitive verbs are often described as “verbs lack[ing] direct objects,” while transitive verbs are described as the ones that do require a direct object (Mahootian:48). Recall that in Persian the subject can be dropped. This feature of Persian can result in potential for ambiguity: in a given situation, how do we know if we are dealing with an SV construction such as (59), or a subject-drop ((S)OV) version of (60)? In other words, the phrase in (61) below is ambiguous:

- (59) æsb mord
 horse die-PAST-3SG
 ‘the/a/some horse died’
- (60) Aræš æsb košt
 Arash horse kill-PAST-3SG
 ‘Arash killed a/some horse(s)’
- (61) æsb košt
 horse kill-PAST-3SG
- (i) ‘[some animate] killed a/some horse(s)’
 (ii) ‘a/some horse killed [some animate]’

Thus, the phrase in (61) can mean a horse had killed an (unspecified) entity, as exemplified in (62) and (63) below. In another situation, however, we may be dealing with a situation in which an (unspecified) entity has killed the horse, as demonstrated in (64). Observe:

- (62) A: æsb či-kar kærd?
 horse what do-PAST-3SG
 ‘What did the horse do?’
- B: æsb košt.
 horse kill-PAST-3SG
 ‘[the] horse killed’

Or, adding more context to our scenario:

- (63) A: aræš čī šod?
 Arash what become-PAST-3SG
 ‘What happened to Arash?’
- B: æsb košt-eš.
 horse kill-PAST-3SG-PRO-3SG
 ‘a/some horse killed him.’

In the two examples above, the horse is the AGENT, doing the killing. Now, compare (B) above to the following:

- (64) A: aræš čī-kar kærd?
 Arash what-work do-PAST-3SG
 ‘What did Arash do?’
- B: æsb košt-eš.
 horse kill-PAST-3SG-PRO-3SG
 ‘[he] killed a/some horse’

Here, ‘Arash’ is the AGENT, and ‘horse’ is the PATIENT. Note, however, that the presence of *ra* alone, disambiguates the phrase. That is, only one reading is possible:

- (65) æsb **ra/ro/o** košt-eš
 horse *ra* kill-PAST-3SG-PRO-3SG)
 ‘[s/he] killed the horse.’

In (65) above, ‘horse’ can only be considered a PATIENT. In contrast, the intransitive verb *mordæn* (‘to die’) does not present an ambiguous situation in that its (one) required term can never get *ra*:

- (66) MORDÆN (æsb)
 die horse
- (67) æsb mord
 horse die-PAST-3SG
 ‘the/a/some horse died’

But not

* æsb **ra** mord

In the examples above, we saw that while there was no confusion in the interpretation of a sentence involving an intransitive verb *mordæn* ('to die'), a transitive verb *koštæn* ('to kill') functioned differently. Of course, these results are consistent with the claim that *ra* marks direct objects, and that it does not occur with subjects.

2.1.3 *ra* and its Referential Status

The claim in the literature is that while *ra* can occur in a variety of contexts, it primarily occurs with direct objects that are definite/specific. Recall that my concern is the referential status of this morpheme, as there is some disagreement in the literature regarding whether *ra* occurs with definite or specific DPs. Previous scholarship on *ra* in Persian has involved the use of terms such as definiteness and specificity, often without precise definitions of what these terms mean or how they relate to one another. In this dissertation, I use the Givenness Hierarchy (Gundel, Hedberg, and Zacharski 1993) in determining the referential status of *ra*-marked DPs. The Givenness Hierarchy makes it possible to use a scalar framework, thus allowing us to resolve the apparent controversy that may arise from the application of dichotomous frameworks traditionally used to explicate the referential functions of *ra*. In Chapter 3 below I will present the methodology used in data selection, followed by Results and Discussion in Chapter 0, where I present a detailed analysis of how the Givenness Hierarchy helps explicate the referential status of *ra*.

In the following chapter, I will present the sources of the data for this study, and the methodology used to determine the cognitive status of nominal expressions marked with *ra*. Note that part of the difficulty in pinpointing the distribution of *ra*, and the

complexity of any subsequent analysis of it, is that Persian speakers have used *ra* in increasingly divergent linguistic environments over time⁴². Moreover, there are many different varieties of Persians in use today, as I indicated in the Background section in Chapter 1. My work on *ra* covers a variety of “Persians” to include spoken Tehrani and several genres in the written dialect.

⁴² Contemporary Persian does contain elements from Middle or Old Persian. Indeed, due to sociolinguistic factors such as standardization, contact, and education, as discussed earlier, contemporary spoken and written Persian often does contain such elements to varying degrees.

3 The Study

In Section 3.1 I describe the issues related to the referential status of *ra*-marked nominal expressions by reviewing the relevant previous research, including the claims about the referential status of *ra*. Further, I describe my study and provide justification for using the Givenness Hierarchy (GH) framework. In reviewing previous research and presenting the contradictory claims about *ra*, I provide a rationale for using the GH instead of the traditional specificity/definiteness dichotomies. Section 0 is dedicated to methodology to include data collection and the procedure used for determining cognitive status of *ra*-marked DPs in this study.

3.1 Cognitive Statuses of *ra*-Marked Expressions in Discourse

3.1.1 Definiteness, Specificity, and *ra*

There is some disagreement in the literature regarding whether *ra* occurs with definite or specific DPs. Lambton (1953) states that “[w]hen a *definite* noun is the direct object of the verb, this is marked by the addition of the suffix [*ra*].” She provides the following example as support:

- (68) **ketab** *ra* be man dad
 book **ra** to l-sg gave-3sg
 ‘s/he gave the book to me’

Phillott (1919) also states that if a noun is definite and “it is in the accusative” it requires *ra*. He gives the following examples (note that Phillott does not give glosses, but I have included them here for clarity):

- (69) ab **ra** bedeh
 water ra give
 ‘Give (me) *the* water’
- (70) ab bedeh
 water give
 ‘Give me water’
- (71) do æsp **ra** did-æm
 two horse ra see-PAST-1SG
 ‘I saw *the* two horses’
- (72) do æsp did-æm
 two horse see-PAST-1SG
 ‘I saw two horses’
- (73) emruz dær bazaar æslæn æsp næ-did-æm
 today in bazaar at all horse NEG-see-PAST-1SG
 ‘Today I saw no horses in the bazaar’

Until Browne (1970), the claim had been that *ra* marks definite DP’s. Browne (1970), to my knowledge, is the first researcher to claim that *ra* marks specificity. His rationale for this claim appears not to be based on a corpus study, but on what is perceived to be a contradiction on a theoretical level: a problem with analyzing *ra* as marking definiteness is that it can co-occur with the suffix *–i*, a morpheme considered by some researchers to mark indefiniteness⁴³. Browne posits that if *ra* marks an NP [sic] that is definite, then it should not mark an NP that is already marked with an indefinite marker. Note that this apparent contradiction is inherent in the dichotomous nature of the definiteness/indefiniteness distinctions⁴⁴. While, as noted above, Phillott (1919) states that *ra* occurs with definite noun phrases and Browne (1970) claims that it occurs with specifics, Ghomeshi (1997) offers the following example to claim that *ra* can

⁴³ The status of *–i* is unresolved, as it is claimed to mark indefiniteness by some researchers, and definiteness by others. Furthermore, if *–i* marks indefiniteness, it is not entirely clear whether it marks indefinite specifics or indefinite non-specifics.

⁴⁴ The Givenness Hierarchy does not present such a contradiction, as statuses on the hierarchy are in an implicational entailment relation and “definiteness”/ “indefiniteness” are not primitive concepts.

optionally co-occur with indefinites. (Note that Ghomeshi’s claim is not inconsistent with Browne in that specifics can be definite or indefinite. See Figure 5 below.)⁴⁵

- (74) ketāb-**i** xarid-am
 book-**INDEF** bought+1sgS
 ‘I bought a book’ [Ghomeshi 1997, (1)(c)]
- (75) ketāb-**i-ro** xarid-am
 book+**INDEF+ra** bought+1sgS
 ‘I bought a (certain/particular) book.’ [Ghomeshi 1997, (1)(d)]

But compare (75) with (76) below, which shows that the *i+ra* combination is not always interpreted as “a (certain/particular)”, as it is sometimes assumed.

- (76) ketab-**i** *o* ke bāram xāride-bud-i xund-æm
 book+**i** *ra*⁴⁶ that for me bought-be-PSPT-2Sg read-1SG
 ‘I read the book that you had bought for me’

In addition, (77) shows that the *i+ra* combination cannot always be analyzed as an inseparable single morpheme as *ra* can occur at the end of a relative clause whose head is marked by *i*. Also, note that regardless of whether *ra* occurs after the head noun or after the relative clause (i.e., the end of the whole DP), ‘ketab-*i ke bāram xāride boodi*’ must be interpreted as definite.

- (77) ketab-**i** ke bāram xāride-bud-i *o* xund-æm
 book+**i** that for me bought-be-PSPT-2Sg *ra* read-1SG
 ‘I read the book that you had bought for me’

⁴⁵ The possibility of both a definite and indefinite interpretation for forms marked with an ‘indefinite’ marker is not inconsistent with accounts such as the Givenness Hierarchy framework of Gundel et al (1993) which views ‘indefinite’ markers as simply unspecified for definiteness (i.e. unique identifiability). Thus, in English, in some contexts where the conceptual content encoded in a phrase uniquely identifies the referent, either a definite or indefinite article is possible with no difference in interpretation.

⁴⁶ ‘o’ here is glossed as *ra* as it is a variant of it (in spoken Persian).

Again, Browne (1970) resolves the apparent contradiction of a proclaimed definiteness marker *ra* co-occurring with an indefinite marker *-i* by proposing that *ra* marks a direct object for specificity rather than for definiteness. Since Browne (1970), other researchers have followed suit. For example, Windfuhr (1990) claims that it is not the case that *ra* marks definiteness; he states that *ra* is the marker of “topicalization or specificity”. Windfuhr gives the following example to demonstrate that *ra* can topicalize adverbials:

- (78) emšab ra inja baš
 Tonight ra here be
 ‘be/stay here (for) tonight’⁴⁷,

Karimi (1990), following Browne, also claims that *ra* is a marker of specificity rather than definiteness. Karimi defines the notion of *specificity* to mean "the selection of a particular individual(s) from a set of individuals." According to Karimi, NPs [sic] fall into two discrete categories, *specific* and *non-specific*. In this model, specifics can further split into two distinct categories of indefinite (known only to the speaker) and *definite* (known to both the speaker and the addressee). Further, non-specifics are considered to be distributed along two discrete lines: *indefinite* and *generic*. Note that while *generics* are always non-specific and *definites* are always specific, *indefinites* are ambiguous with respect to specificity, in that an indefinite can either be specific or non-specific. The diagram below, adapted from Karimi (1990), helps visualize these concepts and their relationships to one another.

⁴⁷ It could be argued that ‘tonight’ is still definite here; so, this example is consistent with the claim that ‘*ra*’ marks definiteness.

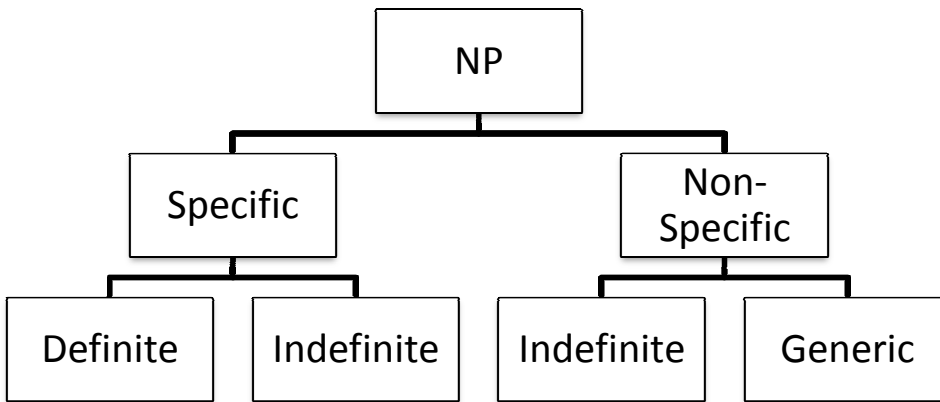


Figure 5 Specificity, Definiteness, and Genericity. (Adapted from Karimi (1990))⁴⁸

Browne (1970), Karimi (1989), and Windfuhr (1990) have disputed the traditional argument that *ra* marks definite direct objects and have claimed that *ra* is primarily a marker of specificity. However, the disagreement has continued, as Ghomeshi (1997) and Mahootian (1997) have disagreed with this analysis and have gone back to the traditional claim. Ghomeshi (1997), while agreeing that a direct object can “optionally” take *ra*, claims that when it does, it signals definiteness. Compare (79) and (80) below.

⁴⁸ The framework depicted in Figure 5 makes it difficult to capture the generalization that generics share properties with definites and are marked as definite in many languages. For example, as can be seen below, ‘dog’ can be interpreted as definite and generic (c.f., Footnote 53):

diruz	daštəm	tu	park	rah	mi-ræft-əm
yesterday	had-1SG	in	park	route	DUR-go-PAST-1SG
ke	yedæfe	sag-i-o	did-əm	ke	dašt
that	suddenly	dog-i-ra	see-PAST-1SG	REL	had
pars	mi-kærd.				
bark	DUR-do-PAST-3SG				

‘Yesterday I was walking in a/the park and suddenly I saw a/the dog that was barking.’

and focused on the way *ra* “signals identifiability of discourse referents,” arguing for an analysis that considers the identifiability phenomenon as an aspect of the theory of information flow to explain the behavior of this morpheme as it is used in actual discourse. Shokouhi and Kipka report that their work suggests that *ra* “is used by the speaker to ‘advise’ the hearer about the identification of referents” and propose that *ra* “signal[s] to the addressee that identification of an object be undertaken.” In their study, Shokouhi and Kipka consider ‘given’ versus ‘new’ information, as well as accessible information. An accessible referent, Shokouhi and Kipka state, is a “referent [that] is either derivable from the situation of the discourse or it is a given referent which has not been activated for some time in the discourse.” They report that 98.7% of the tokens with *ra* are classified as identifiable.

While it is not entirely clear in Shokouhi and Kipka (2003) how “identifiability” is defined or applied to each token, I believe the approach from an information structure perspective is on the right track. I agree with Shokouhi and Kipka that a discourse-based approach is needed (Sadrai 1998, 2003, and 2008), and that a model must be developed in which we move beyond the simple discrete and binary, either/or systems of plus/minus definiteness/specificity. However, while Shokouhi and Kipka talk about moving away from binary models, it appears to me that the Identifiable/Non-identifiable system they adopt is itself a binary system. Recall also that Mahootian states that “*ra* marks object NPs [sic] toward the higher, more definite end of the scale.” While I agree with Mahootian’s observations that a “scalar” approach is needed, it is important to distinguish between the type of ‘gradable’ scalar approach implied here and the ‘entailment’ scale afforded through the use of the Givenness Hierarchy I employ in my

analysis⁵¹. Since my main concern here is the referential status of *ra*, it is imperative that we define concepts such as *specificity* and *definiteness*.

The literature on *ra* often invokes concepts of *specificity* and *definiteness*, but these terms are not always defined as precisely as they could be, and their relation to one another remains unclear. In considering the conditions under which *ra* occurs, I use the Givenness Hierarchy (Gundel, Hedberg, and Zacharski, 1993), a framework that makes explicit the notions of specificity and definiteness and their relations to one another. The Givenness Hierarchy (GH), as we shall see shortly, allows a more precise definition of definiteness and specificity in terms of more primitive concepts that allow for finer distinctions to be made. The framework also regards these notions to be implicationally related and, in that sense, scalar. The main goal of my current work, to study DPs in Persian in naturally occurring data to determine the cognitive status signaled by each *ra*-marked DP using the Givenness Hierarchy, will be further refined in the next section.

⁵¹ It should be noted that Mahootian's work cited here is a text in the Descriptive Grammars series, where only a few pages are dedicated to the treatment of *ra* in a work that is intended to cover most basic aspects of the Persian grammar. We would not expect such a text to deal with issues of theory at the level treated here in this dissertation or in similar works cited here.

3.1.2 The Givenness Hierarchy and the Concepts of Definiteness and Specificity

As mentioned in the previous section, various and sometimes contradictory claims about the functions of *ra* have been made in the literature. The current work addresses the following question: What are the cognitive statuses of referents of nominal expressions marked with *ra*? I have also considered all DP's in the corpus that are marked not only with *ra*, but also with another morpheme, *-i*, whose referential status remains unclear. I will address all the *i+ra* marked DPs in my data separately in Section 4.9.2.

As noted above, the Givenness Hierarchy (Gundel, Hedberg, and Zacharski 1993) makes it possible to define the concepts of specificity and definiteness and their relation to one another more precisely in terms of a set of cognitive statuses (memory and attention states) assumed of the hearer at a given point in the discourse—the cognitive statuses are implicational and scalar by definition. The definitions for each of the Cognitive Statuses are given below.

Cognitive Status:	Definition:
IN FOCUS	The referent is in the addressee's current center of attention, and as such, also activated.
ACTIVATED	The referent is in the addressee's current short term (working) memory, and as such, also familiar.
FAMILIAR	The addressee is already familiar with this object, i.e., has a representation in memory.
UNIQUELY IDENTIFYABLE	The addressee can associate a unique representation with the referent (at least enough to distinguish it from other representations in memory) based on an already existing representation in memory, on unique conceptual content encoded in the phrase itself, or by way of a bridging inference to a recently activated entity.
REFERENTIAL	The speaker is referring to a particular object(s), and the addressee is expected to construct a representation of the referent in question (and as such also the type it belongs to) by the time the utterance is processed in its entirety.
TYPE IDENTIFIABLE	The addressee can identify the type of object being described.

Table 2 Cognitive Statuses and Definitions, provided by Jeanette Gundel in personal communications

Cognitive statuses in GH form an implicational hierarchy, given below along with relevant English forms (for reference) that encode these statuses.

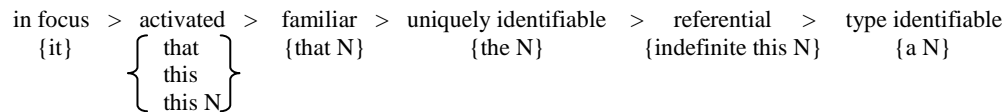


Figure 6 The Givenness Hierarchy (Gundel, Hedberg, Zacharski 1993)

These statuses are encoded by definite and indefinite articles, personal pronouns, demonstrative pronouns, and demonstrative determiners across languages. The Givenness Hierarchy predicts that speakers use a particular form based on their assumptions of the location of a referent in the memory and attention of a hearer at the point in the discourse just before the form is used. The statuses range from least restrictive to most restrictive, as shown below in Figure 7.

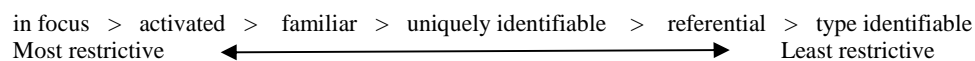


Figure 7 Level of implicational restrictiveness of the Givenness Hierarchy

Further, as noted earlier, these statuses are implicationally related. Each status on the hierarchy entails all lower statuses, by definition, and is therefore included by all lower statuses, but not vice versa (i.e., the entailment is unidirectional).

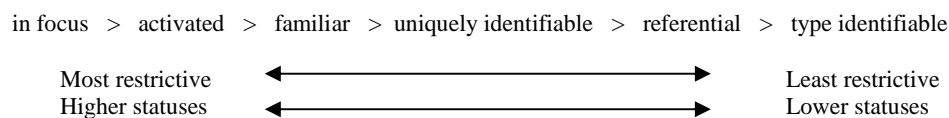


Figure 8 Higher statuses are associated with most restrictive statuses, and lower statuses with the least restrictive

The Venn diagram below visually demonstrates the implicational relationship inherent in the framework.

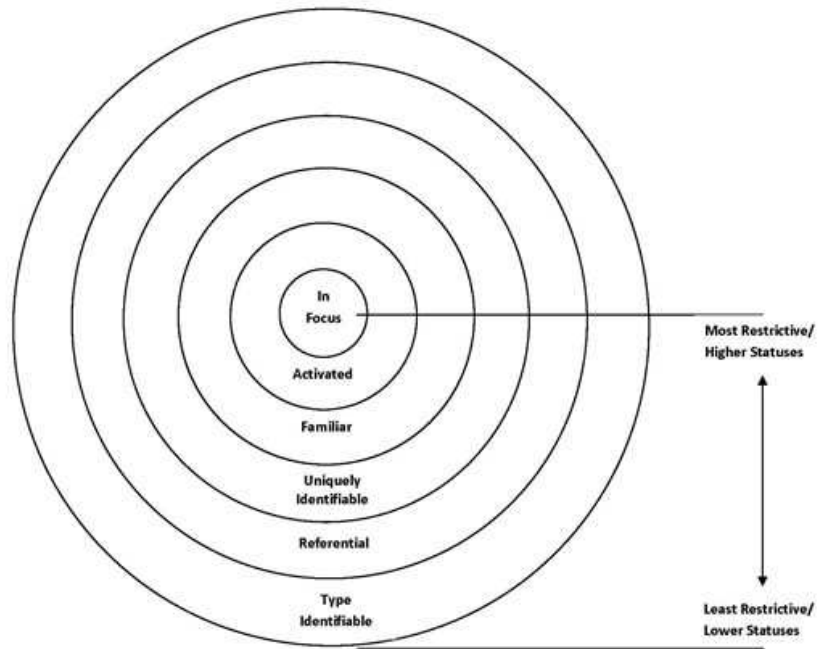


Figure 9 The Cognitive Statuses are Implicationally Related

Therefore, in using a form that explicitly signals some status, the speaker signals that (i) the associated cognitive status is met; and, (ii) all lower statuses have been met⁵². For example, if something is coded by a form that explicitly signals the status *in focus*, it is also, by definition, *activated*, *familiar*, *uniquely identifiable*, *referential*, and *type identifiable*. If something is *referential*, it is also *type identifiable*, and so on.

⁵² In using a particular form, the speaker signals to the addressee the cognitive status (location in memory and attention) of a particular referent; e.g., whether the addressee is already familiar with the referent, is the referent in the current short term memory, or is attention focused on the referent, etc.

Since in signaling one status, a particular form also meets all lower statuses, it follows that, in principle, each form can be used to encode any of the higher cognitive statuses, since these would meet the minimum required status signaled by the form used. For example, since ‘the’ in English signals that something is *uniquely identifiable*, a phrase such as ‘the cat’ (which minimally signals a cat that is uniquely identifiable by the addressee) can be used to refer to a cat that is not only uniquely identifiable, but in the current center of attention (in focus), since for a cat to be in the center of attention, it also has to be *uniquely identifiable* (and *activated* and *familiar*, as well.)

Recall that Cognitive statuses are the location of referents in the memory and attention of a hearer. Figure 10 below is an extension of Figure 5 (presented earlier) that helps visualize how the Givenness Hierarchy maps into the traditional concepts, and extends those notions in terms of more fine-grained and precise cognitive notions.

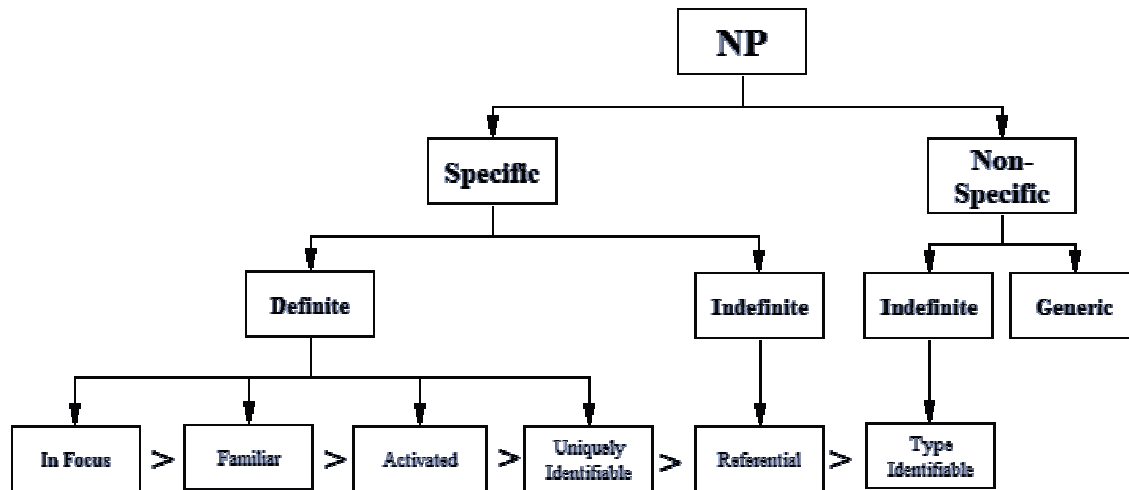


Figure 10 Traditional Concepts Map into GH and are Extended ⁵³

Specificity, then, in this dissertation, will be defined as ‘referential’ or higher, and *definiteness* will be defined as ‘uniquely identifiable’ or higher.

⁵³ Although not represented graphically in this diagram, generics are at least uniquely identifiable; however, what is uniquely identifiable is the type. There is an overlap between generic and uniquely identifiable in that if ‘generic’ is defined as something that refers to a *class* of entities, then it collapses ‘type identifiable’ and ‘uniquely identifiable’. However, while there is a complete overlap between generic and uniquely identifiable, and while all generics are type identifiable, not all type identifiables are generic (i.e., some type identifiables are indefinite since they refer to any member of a class/type)

3.2 Methodology

The goal of this section is to briefly explain the procedures for data collection and analysis. The data for this dissertation was collected over the course of a number of years, and came from several different sources, to ensure not only depth, but also breadth. As mentioned earlier, the purpose for choosing a variety of sources was to ensure that the results reflect language in use in a variety of genres and contexts. Therefore, I examined forty (40) sources of data, including short stories, newspaper articles, children's stories, and a telephone conversation. These sources of data are presented below in Table 3.

Short stories: Four (4) selections from a book called 'Sea of Jewels' (Daryaa-ye Gowhar), including the Prologue and Introduction, in addition to two randomly selected short stories from the text.
Newspaper Articles: Eight (8) articles from the <i>Sports</i> section of <i>Iran</i> Newspaper Three (3) articles from the <i>Culture</i> sections of two different newspapers: two (2) articles from on-line <i>IRNA</i> (Iranian News Agency) and one article from <i>Aftaab-e Yazd</i> ("Yazd's Sunshine"). Nineteen (19) articles from the <i>World</i> sections: one (1) article from on-line <i>IRNA</i> (Iranian News Agency); twelve (12) articles from <i>Keyhaan</i> ("World"); and, six (6) articles from <i>Hamshahri</i> ("Fellow Townsman").
Children's Stories: Five (5) children's stories
Speech Corpus: A fifteen (15) minute telephone conversation between two friends, collected by Linguistic Data Consortium (LDC) in a project called "CallFriend".

Table 3 Sources of Data Collected

The 40 sources of data listed in Table 3 contained a total number of 17,612 words, yielding 408 instances of *ra*-marked DPs. Each phrase containing *ra* was examined and coded for the highest cognitive status of the referent of the nominal expression, using the Givenness Hierarchy (Gundel, Hedberg, and Zacharski 1993). In Section 3.2.1 below, I will briefly outline the overall

approach to data selection, the number of words in each source, as well as the number of *ra*-marked DPs encountered. Each set of data is treated separately, beginning with the speech corpus, short stories texts, the newspaper articles, and the children's stories. Again, the purpose of that section is to present some preliminary observations regarding each source of data presented in Table 3 above. Section 3.2.2 outlines the procedure for determining the highest cognitive status of the referent of the DPs containing *ra*.

3.2.1 Data Selection

Speech Corpus

The speech corpus used for this study was an audio file of a telephone conversation collected by LDC (Linguistic Data Consortium)⁵⁴ as part of a project called *CallFriend*. This CallFriend Persian corpus (conversation number FA_4221) is a 14 minute and 55 seconds conversation that takes place between two female native speakers of Persian. LDC reports that the originator of the call (“caller”) is a 29 year-old female with 20 years of education who grew up in Tehran. The only information provided about the “callee” is that she is a female. Based on pronunciation and vocabulary choice, I would say that the callee speaks the Tehrani dialect as well. LDC also reports that participants were both aware of being recorded, but that they were not given any guidelines concerning what they should talk about.

In preparation for examining this recorded speech, I first transcribed the data. For doing so, I used a digital media playback program called Winamp3. This particular program was chosen from among 4 softwares considered, because of its ability to fast-forward and rewind through the audio file at intervals of 5 seconds per click. Although, ideally, a program with shorter intervals (i.e., fractions of a second) would have worked better, Winamp3 provided the shortest play-back intervals of the 4 softwares examined for this purpose. Once the conversation was transcribed on paper, I transliterated the transcribed Persian text, using a modified (North American) version of the International Phonetic

⁵⁴ The Linguistic Data Consortium is an open consortium of research organizations, created to address data shortage for research and development. For more information please visit their website at <https://www.ldc.upenn.edu/>

Alphabet, as used by linguists in North America. During the 3rd step in the process, I glossed and translated the entire text.

The dialogue consisted of 350 turns between the two interlocutors; the total word count was 2,266 words which included 37 *ra*-marked DP's.

Written Data

Short stories

Two stories were selected randomly from a collection that included 52 short stories. The collection called *dærya-ye gowhær* ('Sea of Jewels') was published in 1988. In addition to the two short stories from this collection, *širin kolah* ('Sweet Hat') and *fæth-e rumi* ('Victorious Roman'), the Prologue and the Introduction to this compilation were examined as well. Table 4 below gives an overview of the findings.

	Word Count	<i>ra</i> Count
Prologue	460	6
Introduction	576	11
Short Story: 'Sweet Hat' ("širin kola")	2505	89
Short Story: 'Victorious Roman' ("fæth-e rumi")	2005	61
Totals:	5,546	167

Table 4 Total Number of Words in the Data and Number of Occurrences of *ra*

A total of 5,546 words were examined and 167 occurrences of *ra* were identified. Again, further details of the findings will be presented in Section 3.1

Newspaper Articles: ‘Sports’ Section

A total of eight (8) articles from the ‘Sports’ Section of a newspaper called “Iran” were examined (Iran: March 19, 2003). These articles, together, consisted of 2,898 words, consisting of 61 *ra*-marked phrases which were subsequently coded for the highest cognitive status. Table 5 below outlines the details.

	Word Count	<i>ra</i> Count
‘Nekunaam: Mikhaahim Baa bacheh-haa-ye Paas be Shomaal Beravim’ (“Nekunam: We Want to Go Up North with the <i>Pas</i> Players”)	174	4
‘Payaam-e Peykaan ham Ta’m-e Shekast ra Cheshid’ (“ <i>Payam-e Peykan</i> Also Tasted the Flavor of Defeat”)	82	2
‘Tahavvol dar Fa’aaliyat-haa-ye Varzeshi-e Mantagheh’ (“Change in Sports’ Activities in the Region”)	122	1
‘Perspolis Choob-e Kharid-haa-yash raa khord’ (“ <i>Persepolis</i> Was Punished Due to its Purchases [of new players]”)	950	23
‘Fulaad be Bazikonaanash Eidi Daad’ (“ <i>Fulad</i> ” Gave its Players [New Year] Gifts”)	133	0
‘Aaghaazi Bad, Paayaani Badtar’ (“A Bad Beginning, A Worse Ending”)	943	20
‘500 Hezar Dollar baraa-ye Seh Maah’ (“500 Dollars for Three Months”)	265	7
‘Estili: Faghat Pahlevaan-e Zende haa Eshgh Ast’ (“Estili: <i>Hurray</i> Only for the Alive Hero”)	229	4
Totals	2,898	61

Table 5 Eight Articles from the Sports Section of ‘Iran’ Newspaper

Newspaper Articles: ‘Culture’ Section

Three articles from the ‘Culture’ section of two different news sources were examined.

Two of the articles came from *IRNA* (Iranian News Agency) and one came from *Aftab-e Yazd* (“Yazd’s Sunshine”). A total of ten (10) *ra*-marked DP’s were identified from among 880 words. Details are provided in the table below.

	Date Published	Word Count	ra Count
On-line <i>IRNA</i> (Iranian News Agency):	Nov. 5, 2003		
‘Daaneshgaah-haa-ye Keshvar-haa-ye Saheli-e Khazar Tafaahom-naameh-ye Hamkaari Emzaa Kardand’ (“The Universities of the Countries Bordering The Caspian [Sea] Signed a letter of Mutual Understanding of Collaboration”)		150	0
‘Daryaa-ye Khazar Faghat baraa-ye 100 Saal Markaz-e Tavajjoh va Alaaghe-ye Afkaar-e Omumi Khaahad Bud’ (“The Caspian Sea Will Only Be at the Center of Public Attention and Interest for 100 years”)		405	3
<i>Aftaab-e Yazd</i> (“Yazd’s Sunshine”):	Nov. 6, 2003		
‘Abutoraab Khosravi: Djamaalzaadeh Zabaan-e Kukhe va Baazar raa be Adabiyyaat-e Daastaani Tabdil Kard’ (“Abutorab Khosravi: Djamalzadeh Transformed the Colloquial Language to Literary Fiction”)		325	7
	Totals	880	10

Table 6 Three Articles from 'Culture' Sections of IRNA and Aftab-e Yazd

Newspaper Articles: ‘World’ Section

A total of nineteen (19) articles from the ‘World’ sections of three news sources were selected at random, yielding 4,343 words and eighty five (85) *ra*-marked DPs. One article from the online news source *Iranian News Agency* (March 24, 2003) yielded two (2) *ra*-marked DPs in a data set consisting of 205 words:

	Word Count	ra Count
‘Yek Ma’aareez-e Araaghi az Taghiir-e Barnaame-ye Jang-e Amrika Alaihe Araagh Sokhan Goft’ (“An Iraqi Opponent of The American Change of Plans for War Against Iraq Spoke”)	205	2

Table 7 One Article from 'World' Section of IRNA

Twelve articles were selected at random from the ‘World’ Section of March 19, 2003

Keyhan (“Universe”), which yielded 41 *ra*-marked DPs in a data set consisting of 2,140.

	Word Count	<i>ra</i> Count
‘Towte’e-ye Kharaabkari-e Emrikaa-i-haa dar Kubaa Naakaam Mand’ (“The American Plot for Destruction in Cuba Remained Unsuccessful”)	123	4
‘Fard-e Mosallah-e Yamani 2 Karshenaas-e Emrikaai va Kaanaadai raa be Halaakat Resaand’ (“Armed Yemeni Killed two American and Canadian Experts”)	101	2
‘Tashdid-e Tadaabir-e Aminyati Atraaf-e Markaz-e Siyaasi-Nezaami-e Emrikaa dar Daakhel va Khaarej-e Keshvar’ (“Repetition of Security Policy Surrounding the Internal and External American Political-Military Center”)	264	2
‘E’teraaz-haa-ye Jahaani Alay-he Jang Sheddāt Gereft’ (“World Objections to Against War Intensified”)	322	4
‘Neyoyork Zir-e Kontorol-e Niruhaaye Zedd-e-shoresh Gharaar Gereft’ (“New York Was Placed Under the Control of Anti-riot Forces”)	192	3
‘Saddam va Raamsfeld 18 saal ba’d’ (“Saddam and Rumsfeld, 18 Years Later”)	113	2
‘Emrika: Agar Saddam Ham Beravad Hamleh Mikonim’ (“America: We Will Attack, Even if Saddam Leaves”)	169	11
‘Lomond: Behbud-e Chehreh-ye Amrikaa dar Jahaan Gheyr-e-momken Ast’ (“Le Monde: Improving America’s Reputation in the World Is Impossible”)	121	1
‘Dokhtar-e Saalvaador Alandeh Re’is-e Majles-e Nemaayandegaan-e Shili Shod’ (“Salvador Allende’s Daughter Elected as Speaker of Chile’s House”)	177	1
‘Vakil-e Mohjebbeh-ye Faraansavi raa be Maraasem-e Tahlif-e Vokalaa Raah Nadaadand’ (“They Did Not Allow Covered [with Islamic Hejab] French Lawyer to the Lawyers’ Oath Ceremony”)	104	3
‘Az Tars-e Marg Mordand’ (“They Died of Fear of Death”)	130	3
‘Hemaaghat-e Digar-e Bush’ (“Bush’s Other Foolishness”)	324	5
Totals	2,140	41

Table 8 Twelve Articles from 'World' Section in *Keyhan*

Also, six randomly selected articles from the March 19, 2003 publication of *Hamshāhri* (“Fellow Townsman”) yielded 1,998 words and 42 *ra*-marked DPs.

	Word Count	ra Count
‘Mozaakeraat-e Iran va Yaman dar baareh-ye Bohraan-e Araagh’ (“Discussions of Iran and Yemen about the Iraqi Crisis”).	360	8
‘Baaztaab-e Jahaani-e oltimaatom-e Bush be Saddam’ (“World’s Reaction to Bush’s Ultimatum to Saddam”)	615	18
‘Sherkat-haa-ye Nafti-e Jahaan Kharid-e Naft-e Araagh ra Motafaghef Kardand’ (“Oil Companies in the World Have Stopped the Purchase of Iraqi Oil”)	252	5
‘Tasvib-e Ghanoon-e Ekhtiyar-aat-e Nakhost Vazir-e Felestin’ (“Ratification of the Authority Law of the Palestinian Prime Minister”)	546	10
‘Dar baareh-ye Esteghraar-e Niruhā-ye Emrikaai Parleman-e Torkiyeh baar- e digar Tasmim Migirad’ (“The Turkish Parliament once again Makes a Decision about the Settlement of the American Forces”)	152	1
‘Kashf-e Selaah va Mavaadd-e Monfajereh dar Riaaz’ (“Discovery of Weapons and Explosive Materials in Riyadh”)	73	0
Totals	1,998	42

Table 9 Six Articles from 'World' Section in *Hamshahri*

Children's Stories

Further, a set of five children's stories were examined, yielding 48 *ra*-marked DPs in a data set containing 1,679 words. These children's stories were all published in 1953 in Tehran.

	Word Count	<i>ra</i> Count
'Olaagh va Baar-e Namak' ("Donkey and The Load of Salt")	360	8
'Gorg va Barreh' ("Wolf and Lamb")	380	13
'Khar dar Lebaas-e Shir' ("Donkey in Lion's Clothes")	323	10
'Shekam-e Ghurbaaghe' ("Frog's Stomach")	256	11
'Murcheh va Parvaaneh' ("Ant and Butterfly")	360	6
Totals	1,679	48

Table 10 Five Children's Stories

Aggregated results for all data sets are provided in Appendix A.

Constructed Examples and Grammaticality Judgments

In addition to naturally occurring data, I have supplemented my study with constructed examples. In determining the syntactic and pragmatic acceptability of the naturally occurring data, as well as that of constructed sentences in this study, grammaticality was determined based on my own native intuition, as well as that of six (6) other native speakers in informal grammaticality judgment exercises. I am a native speaker of Persian with 11 years of formal education in Iran. I also have native ability in Tehrani, Golpayegani, and some knowledge of the Esfahani dialects⁵⁵. In addition to my own native judgments, at various stages of this study, at various stages of development, I have

⁵⁵ I have lived and have gone to school in Golpayegan (a city in the central-western part of Iran, with its own dialect), and in Tehran. Further, my family on my mother's side is from Esfahan, a city in central Iran with its own distinctive dialect, where I spent many summers staying with relatives and attending short courses at The Iran-America Institute ('Anjoman-e Iran va Emrika') in late 1960's and in 1970's.

consulted one or more other native speakers of Persian. One speaker is a 70 year-old speaker of Persian with a 9th grade education in Iran. She was born in the city of Esfahan and grew up in that city until the age of 8, before moving to a small village just outside the city of Golpayegan, nestled in the Zagros mountains in the western-central part of Iran. The second speaker is a 35 year-old with an MD from University of Tehran. She is bi-dialectal in Golpayegani and Tehrani. My third consultant is a native speaker of Tehrani with a 12th grade education in Iran, and a BA from University of Minnesota. The fourth consultant is a Tehrani speaker, with a BA from Iran, and a Masters degree from University of St. Thomas. The fifth speaker is also a Tehrani speaker, with a nursing degree from Iran, and a BA from University of Minnesota. The last speaker is also a native speaker of Tehrani with 10 years of education in Iran, and an MS from University of Minnesota.

3.2.2 Procedure for Determining Cognitive Status of *ra*-marked DPs

Using the Givenness Hierarchy (Gundel, Hedberg, and Zacharski (1993), I first located all *ra*-marked DPs. Next, the intended referent of each DP was identified by searching through the text to find its antecedent. The task was to code the referent of each DP for the highest cognitive status: the location of a referent in the memory and attention of a hearer at the point in the discourse just before *ra* was used. Definitions of cognitive statuses, given in Table 2 above, are repeated below for convenience.

Cognitive Status:	Definition:
IN FOCUS	The referent is in the addressee's current center of attention, and as such, also activated.
ACTIVATED	The referent is in the addressee's current short term (working) memory, and as such, also familiar.
FAMILIAR	The addressee is already familiar with this object, i.e., has a representation in memory.
UNIQUELY IDENTIFIABLE	The addressee can associate a unique representation with the referent (at least enough to distinguish it from other representations in memory) based on an already existing representation in memory, on unique conceptual content encoded in the phrase itself, or by way of a bridging inference to a recently activated entity.
REFERENTIAL	The speaker is referring to a particular object(s), and the addressee is expected to construct a representation of the referent in question (and as such also the type it belongs to) by the time the utterance is processed in its entirety.
TYPE IDENTIFIABLE	The addressee can identify the type of object being described.

Table 11 Cognitive Statuses and Definitions, provided by Jeanette Gundel in personal communications (2010)

To determine the highest cognitive status of the referent in question, the guidelines in the table below were used.

Cognitive Status:	Guidelines for Determining Highest Status:
IN FOCUS	<ul style="list-style-type: none"> the referent was mentioned in a syntactically prominent position (e.g., main clause subject) in the immediately preceding sentence the referent was a higher-level topic that was part of the interpretation of the immediately preceding sentence (whether it had been explicitly mentioned or not) the referent was mentioned earlier in the same sentence
ACTIVATED	<ul style="list-style-type: none"> the referent was mentioned in the previous two sentences the referent was present in the immediate, non-linguistic spatio-temporal (i.e., physical) context the referent was mentioned three sentences previously, but not necessarily since then
FAMILIAR	<ul style="list-style-type: none"> the referent was mentioned at any time previously in the discourse the referent was assumed to be shared knowledge between speaker and hearer (either through shared cultural knowledge or shared experience)
UNIQUELY IDENTIFIABLE	<ul style="list-style-type: none"> the referring form contained adequate descriptive content to construct a new unique representation the hearer was able to identify a unique referent by linking it indirectly to a recently activated referent
REFERENTIAL	<ul style="list-style-type: none"> the referent was mentioned subsequently in the discourse no subsequent references to the entity occur, but it is reasonably clear from the context that the speaker intends to refer to a particular object
TYPE IDENTIFIABLE	<ul style="list-style-type: none"> the hearer can identify the type of thing described, i.e., the hearer can be assumed to understand the meaning of the words and the sense of the expression in general

Table 12 Guidelines for determining cognitive status, provided by Jeanette Gundel

Coding for Cognitive Status of a DP, as a starting point, I first began by considering if the referent of the DP was Familiar. If the referent of the DP was Familiar, I would check to see if it were also Activated or In Focus, in that order; if not, I would go on to consider lower statuses: Uniquely Identifiable, Referential, and Type Identifiable, in that order. If

there was a doubt, I went with the more conservative one; for example, as demonstrated in the example below, if the referent of a DP appeared to be either In Focus or Activated, I chose Activated.

- (81) xob, ma juya-ye hal-et bud-im.
 good 1PL seek-LNKPT health-2SG was-1PL
 ‘Well, we had sought the state of your health’ [or, “we wondered how you were”]
- æz Iran soraq-e -to dašt-im.
 from Iran seek-LNKPT-2SG have-PAST-1PL
 ‘From Iran we had sought of you’
- bær-gæštæn-æm soraq-e-to dasht-æm.
 back-turn[INFINITIVE]-also seek-LNKPT-2SG have-PAST-1SG
 ‘Coming back, we’d asked about you as well.’
- ba šæmsi xanum umæd-in. **xæbær-a-ro**
 with Shamsi Ms. come-PAST-2PL news-PL-ra
- avord-æn.
 bring-PAST-3PL
- ‘You came with Ms. Shamsi. They brought the news.’

In the example above, it is somewhat difficult to determine if *xæbær* (‘news’) is referring to *ba šæmsi xanum umæd-in* (‘you came with Ms. Shamsi’), which would be in the addressee’s current center of attention, and therefore In Focus, or if it is referring to *hal-et* (‘the addressee’s state of health’), which would be considered to be in the addressee’s current short term memory (Activated), at the point in which *xæbær* (‘news’) is uttered. In this case, as stated above, I considered the referent of *xæbær* to be the more conservative “Activated” status.

In the Results section in Chapter 4, I will present my observations of the cognitive status of *ra*-marked expressions for each source of data. I will further report on the most common, as well as ideal examples of each cognitive status represented. I will then come back to the concepts of specificity and definiteness and evaluate how my findings, using

the Givenness Hierarchy, interacts with these traditional concepts and which of the previous claims in the literature it supports.

4 Results

In the Results Chapter I (i) present the Cognitive Statuses of the referents of *ra*-marked expressions found in my data, (ii) discuss the most common Statuses observed, and (iii) provide a representative selection of most common Cognitive Statuses found. Finally, I summarize my findings, describing the patterns I observed: Cognitive Statuses most commonly associated with the referents of the nominal expressions studied.

Table 13 below summarizes the data examined for this study. The data consisted of 17,612 words and included 407 occurrences of *ra*-marked DPs. As reflected in this Table, 15 *ra*-marked DPs were not coded (because they occurred out of context, including quotations, formulaic expressions, and in bulleted items). Also, 14 of the of *ra*-marked DPs were separated from the rest of the corpus and were considered separately, as they co-occurred with the morpheme *-i* (*-i+ ra*). The total number of *ra*-marked DPs coded for cognitive status were 378, of which 119 were coded as In Focus, 57 were coded as Activated, 45 were coded as Familiar, and 157 were coded as Uniquely Identifiable. Examples and discussion of most common cognitive statuses follow Table 13. The *ra*-marked DPs which co-occurred *-i* will be discussed separately, in 4.9.2.

	Word Count	<i>ra</i> Count	Not Coded	In Focus	Activated	Familiar	Uniquely Identifiable	-i+ra
Short stories (including 'Prologue' and 'Introduction')	5,546	167	2	51	25	15	70	4
Children's stories	1,679	48	1	14	9	7	13	4
Newspaper articles: Sports section	2,898	61	3	16		9	33	
Newspaper articles: World section	4,343	84	4	30	11	9	29	1
Newspaper articles: Culture section	880	10	2		1		7	
Speech data	2,266	37	3	8	11	5	5	5
Totals	17,612	407	15	119	57	45	157	14

Table 13 Summary Results of Highest Cognitive Statuses of ra-Marked DPs in the Entire Corpus Examined

4.1 Short Stories: Prologue

The Prologue section of a book of short stories, called ‘Sea of Jewels’ (“dārya-ye gowhær”) that was examined consisted of 460 words, and six (6) DPs were found to contain *ra*. Of the six DPs, one was marked with both *-i* and *ra*, which, as mentioned earlier, will be discussed in 4.9.2. Further, one other DP was excluded because it occurred in a quotation. Note that previous mention of an intended referent is one basis for determining the cognitive status of an intended referent. For this reason, to determine the cognitive status of the referent of a DP, it is imperative to consider an entire text to establish whether a referent has been previously mentioned anywhere in the discourse. As such, a quotation cannot be effectively examined since it has been taken out of its original context. I examined the remaining five DPs for the cognitive status of their referent, and determined that two could be assumed to be at most In Focus, one was coded as Activated, one was coded as Familiar, and one was coded Type Identifiable. When coding each DP, I initially looked to see if the intended referent could be assumed to be Familiar: (i) if the referent was assumed to be shared knowledge; or, (ii) the referent was mentioned at any time previously in the discourse. To determine if the referent was mentioned previously in the discourse, I began with the immediate preceding environment, looking for linguistically introduced forms to identify the intended referent. For example, if the intended referent was mentioned in the same sentence, I coded the referent of that DP as being In Focus. If it was mentioned in the previous two sentences, but not since then, I classified it as Activated. If the referent was mentioned at any time previously in the discourse, I coded it as Familiar. Table 14 below summarizes the findings. Following the Table below, each of the cognitive statuses are presented.

In Focus	Activated	Familiar	Uniquely Identifiable	-i+ra	Not Coded
2	1	1		1	1

Table 14 Highest Cognitive Statuses of ra-Marked DPs in Prologue

In Focus. In (82) below, *an-ha* (them) refers to *□eld-e čaharom væ pæn□om-e mæ□mu□e-ye dærya-ye gowhær* (‘fourth and fifth volumes of the collection of Daryaye Gowhar’) which is mentioned in a prominent (subject position) in the main clause, and as such, is coded as In Focus since it is assumed to be in the addressee’s current center of attention.

(82)	□eld-e volume-LNKPT	čaharom fourth	væ and	pæn□om-e fifth-LNKPT	mæ□mu□e-ye collection-LNKPT
	dærya-ye sea-LNKPT	gowhær jewel	tæqribæn almost	amadeh ready	æst be-Pres-3Sg
	ke REL	enšaëllah God willing	ba with	komæk-e help-LNKPT	aqaye Mr.
				nušyar Nushyar	hæmidi Hamidi
	færzænd-e offspring-LNKPT	an	mærhum that	be-tævan-æm deceased	SUBJ-able-1Sg
	an-ha ra that-PL ra	bæraye for	čap print	amadeh ready	kon-æm. SUBJ-do-1SG

‘The fourth and the fifth volumes of the collection of Daryaye Gowhar are almost ready which with the help of Mr. Nushyar Hamidi and the offspring of that deceased [the author’s son] I may prepare **them** for print.’

In Focus and Activated. In the following data, we will see two instances of *ra*, one considered to be In Focus, while the second one is determined to be Activated.

- (83)
- | | | | | | | | | |
|--|-----------|-------------------------|-------------------|-----------|------------|-----------|-------------|--|
| ostad | doktor | hæmidi | æz | in | se | ʔælamæt | | |
| master | doctor | Hamidi | from | this | three | symbol | | |
| “+” | væ | “+x” | væ | “-“ | bəraye | tæshil | dær | |
| + | and | +x | and | - | for | ease | in | |
| kar | ya | be-gofte-ye | | | xiš | bəraye | penhan | |
| work | or | in-say-LNKPT | | self | for | hide | | |
| negah-daštæn-e | | ʔæqide-æš | | | | dær | bare-ye | |
| look-have-LNKPT | | opinion-self-3SG-POSS | | | | in | about-LNKPT | |
| æsær-i ⁵⁶ | | ke | mi-xand | | | estefade | | |
| work-i | | REL | DUR-read-PAST-3SG | | | use | | |
| mi-kærd, | | | | | | | | |
| DUR-do-PAST-3SG | | | | | | | | |
| ‘The Master Dr. Hamidi, to make it easier to work, used these three symbols, ‘+’, ‘+x’, and ‘-’, in his own words, to hide his opinion about a/the work that he was reading’ | | | | | | | | |
| bæʔdæn | | ke | qərar-šod | | mæn | dær | in-kar | |
| later | REL | agreement-become-PAST I | | | in | this-work | | |
| ʔu | ra | yari | dəhæm | in | raz | ra | | |
| he | ra | help | give-1SG | this | secret | ra | | |
| bəra-y-æm | | baz-gu | | kærd | | | | |
| for-LNKPT-1SG | again-say | | do-PAST-3SG | | | | | |
| ‘Later, when it was agreed upon that I would help him in this work, he recounted this secret to me.’ | | | | | | | | |

In (83) *ʔu* (‘he’) in the second sentence refers to ‘Hamidi’, which is mentioned in the subject of the main clause (i.e., mentioned in a syntactically prominent position), of the previous sentence, and as such, meets one of the criterion for being In Focus. The referent

⁵⁶ Recall that the status of the particle *-i* is unresolved as it has been variously classified as a definiteness or indefiniteness marker. In this work, I remain natural as to the status of this morpheme. Note that here in this example, it can be argued that the referent of ‘æsær’ is Type Identifiable at the point at which it is uttered. However, the occurrence of the relative clause that follows renders it Referential, as the addressee is able to construct a representation of the referent by the time the relative clause that follows is processed. Future work will consider this morpheme, using the Givenness Hierarchy framework.

of *raz* ('secret') can be argued to be Activated. The referent of 'secret' is the 'usage of certain symbols to make it easier to work, and to keep his opinions hidden about the work he was reading' which was mentioned in the previous sentence and therefore the addressee can be expected to have a recently activated referent for 'secret' in mind. That is, the addressee can pick out a referent that is in her/his current working (short term) memory. *raz* ('secret') is assumed to be at most activated, since it does not meet any of the criterion for In Focus (it was not mentioned in a syntactically prominent position, it was not mentioned earlier in the same sentence, etc.).

Familiar. Another instance of *ra*, as mentioned above, co-occurred with a DP that was coded as Familiar. In the next paragraph (following the assertion in (83) above), the author continues to talk about *Hamidi's opinions* about the works and the authors he was reading, and having been asked to publish Hamidi's works.

- (84) ... səʔi xahæm kærd **næzærie-ha-ye**
 Try want-FUTURE do-PAST-3SG **opinion-PL-LNKPT**
- išan** **ra** ta sær-hædd-e emkan
 he-PL-POLITE **ra** upto head-limit-LNKPT possible
- dær kar-e čapp-e
 in work-LNKPT publish-LNKPT
- asar-e ʔu eʔmal konæm
 work-PL-LNKPT 3-SG apply do-1SG
- 'I will try to apply [i.e., convey] **his** (honorable, Hamidi) **opinions**, to the extent possible, in publishing his works.'

The referent of *išan* ('he') is Hamidi. As Hamidi has not been mentioned since the previous paragraph (as the discussion has been about the works and authors he was working with), it cannot necessarily be expected to be in the addressee's working memory; however, it can be assumed to be familiar (i.e., in memory).

4.2 Short Stories: Introduction

In addition to the Prologue section of the book ‘Sea of Jewels’, the Introduction was also examined for occurrences of *ra*-marked DPs. This section consisted of 576 words, and eleven (11) DPs were found to contain *ra*. Of the 11 DPs in this data set, three were coded In Focus, two were Activated, five were coded as Uniquely Identifiable, and one was not coded for cognitive status. Table 15 below summarizes these findings.

In Focus	Activated	Uniquely Identifiable	Not Coded
3	2	5	1

Table 15 Highest Cognitive Statuses of *ra*-Marked DPs in Introduction

Below I will discuss at least one example from each category from Table 15.

In Focus. There were three *ra*-marked DPs in this data deemed In Focus. Here, for brevity, I will present one example below.

- (85)
- | | | | | | | |
|-----------|------------------|------------------|------------|---------------|---------------------|----------------|
| mæn | bær-ru-ye | mæjmuʃ-e | an | bærg-ha | | |
| I | on-on-LNKPT | collection-LNKPT | that | leaf-PL | | |
| dær | ʃalæm-e | xiyal | nam-e | | | |
| in | world-LNKPT | imagination | name-LNKPT | | | |
| “dærya-ye | | gowhær” | gozašt-e | bud-æm | | |
| sea-LNKPT | | jewel | place-PSPT | be-PAST-1SG | | |
| væ | in | ketab | ke | avvalin | bæxš-i | æst |
| and | this | book | REL | first | part-i | be-PRESENT-3SG |
| ke | æz | an | owraq | ʃoda | mi-šæv-æd | |
| REL | from | that | page-PL | separate | DUR-become-PRES-3SG | |
| niz | hæm-in | nam | ra | xah-æd | | dašt. |
| also | also-this | name | ra | want-PRES-3SG | | have-PAST-3SG |

‘In [my] world of imagination I had [bestowed] the name *Sea of Jewels* on that collection of pages, and this book, which is a part of those pages, will also have **this same name** [i.e., *Sea of Jewels*].’

The second occurrence of *nam* (‘name’) here is marked with *ra*, both referring to ‘Sea of Jewels’. Since ‘Sea of Jewels’ was mentioned in the first clause of the sentence, the referent can be assumed to be in the addressee’s current center of attention, and as such, it was coded as In Focus.

Activated. The referent of the *ra*-marked DP in (86) *an ra* (‘that’) in the first sentence, which refers to ‘the book’ mentioned previously in the same sentence, is In Focus.

However, in (87), the referent of *an arezu* (‘that wish’)—the wish to put together a collection of short stories—mentioned in the previous sentence, is coded as at most Activated: it can be assumed to be in the addressee’s current working memory, but not necessarily in focus.

(86)	in this	ketab book	ba with	ketab-ha-i book-PL-i	ke REL	donbal after	dar-æd, have/PRES-3SG		
	mowlud-e birth/ACC-LNKPT			arezu-ha-i wish-PL-i	æst be-PRES-3SG		ke REL	æz from	sal-ha year-PL
	piš prior	dær in	xial-e imagination-LNKPT			mæn 1SG	xane house		
	kærd-e do-PAST-3SG-PSPT			bud be-PAST-3SG	væ and	mæn 1SG	peyvæste perpetually		
	an that	ra ra	ba with	xod self	bozorg big	mi-kærd-æm. DUR-do/PAST-1SG			

‘This book, [along] with other books that follow, is the inception of wishes that had made home in my imagination for years, and I was perpetually cultivating **that** [i.e., the book] with myself.’

- (87)
- | | | | | | | | |
|------------|---------------------|------------|-----|-------------------|--------------|-----------|-----|
| ruh-e | | mæn | ke | setayeš-gær-e | | | |
| soul-LNKPT | | 1SG | REL | worship-NOM-LNKPT | | | |
| honær | æst | tox-m-e | | an | arezu | ra | dær |
| art | be/PRES/3SG | seed-LNKPT | | that | wish | ra | in |
| dæmaq-æm | kašt-e | | | bud ... | | | |
| nose-1SG | plant-PAST-3SG-PSPT | | | be-PAST-3SG | | | |
- ‘My soul, which is a worshiper of art, had planted **the seed of that wish** (i.e., ‘dream’, or ‘idea’) in my mind...’⁵⁷

Uniquely Identifiable. As can be seen in Table 15 above, the majority of the referents of the *ra*-marked DPs in the Introduction to the short stories were Uniquely Identifiable. Of the four *ra*-marked DPs in this category, the following is selected as an example.

- (88)
- | | | | | | | |
|------------|-------------|-----------------|-----------|-------------------------|-----------------|-----------------|
| mæn | xiyal | mi-kon-æm | | ba | entešar-e | in |
| I | imagine | DUR-do-PRES-1SG | | with | publish-LNKPT | this |
| ketab | momken | æst | | šagerd-an-e | | hušyar |
| book | possible | be-PRES-3SG | | student-PL-LNKPT | | astute |
| væ | ba | ferasæt | ra | dær | mæktæb-ha-ye | moxtælef-e |
| and | with | insight | ra | in | school-PL-LNKPT | different-LNKPT |
| næsr | gærdeš | dad. | | | | |
| prose | stroll | give-PRES-3SG | | | | |
- ‘I imagine, with the publication of this book, it is possible to walk the **astute and insightful students** through the various schools of prose.’

In the above example, the referring form *šagerd-an-e hušyar væ ba ferasæt* (‘astute and insightful students’) contains enough descriptive content to pick out not just any students, but those that are considered ‘astute and insightful’; however, there is no reason to assume the referent is already in the addressee’s memory. As such, the referent of this DP is coded as Uniquely Identifiable.

⁵⁷ Note that the author here is exercising poetic license, using ‘nose’ instead of ‘mind’ where the seed of an idea is planted.

4.3 Short Stories—‘Sweet Hat’ and ‘Victorious Roman’

In addition to the prologue and the introduction from the Sea of Jewels collection, two short stories were selected at random. The two collections included a total of 149 *ra*-marked DPs in a data consisting of 4510 words. Table 16 below shows these results.

	Word Count	<i>ra</i> Count
‘Sweet Hat’ (“širin kola”)	2505	89
‘Victorious Roman’ (“fateh-e rumi”)	2005	61
Totals	4510	150

Table 16 Number of words and *ra*-marked DPs in Sweet Hat and Victorious Roman

The results of detailed examination of the referents of each *ra*-marked DP are given in Table 17 below.

	In Focus	Activated	Familiar	Uniquely Identifiable	-i+ <i>ra</i>
‘Sweet Hat’	30	17	9	32	1
‘Victorious Roman’	16	5	5	33	2
Totals	46	22	14	65	3

Table 17 Highest Cognitive Statuses of *ra*-Marked DPs in ‘Sweet Hat’ and ‘Victorious Roman’

Following are some representative examples of the Cognitive Statuses reported in Table 17.

In Focus. In the ‘Victorious Roman’, in one passage there is a reference to the Roman victory, referred to as the *hadese* (‘incident’):

(89)	bæray-e for-LNKPT	mærdom-e people-LNKPT	nazok-bin narrow-see	mowzu□-e topic-LNKPT	deqqæt scrutiny
	væ and	goft-o-gu-i say-and-say-i	peyda find	šod become-PAST-3SG	væ and
	jæm□iyæt-e population-LNKPT	□oræfa intellectuals	moddæt-ha period-PL	æz from	
	baz-gu-ye again-say-LNKPT	hadese incident	gærm warm	bud. is-PAST-3SG	
	hær each	kæs person	væq□e incident	ra ra	be-now□-i in-kind-i
					tæ□bir interpret
	mi-kærd. DUR-do-PAST-3SG				

‘For the astute people, a subject for conversation and scrutiny was discovered and the intellectual circles were warmed with retelling the incident. Each person would interpret the **incident** in [a particular] way.’

In the above passage, in the first sentence a reference is made to an incident, clear from the context that the ‘incident’ refers to the victory of the Romans. Note that this occurrence of *hadese* (‘incident’) [in the phrase, *æz baz-gu-ye hadese* (‘from retelling the incident’)] is not an obligatory constituent, and does not co-occur with *ra*. Further, *væq□e* (‘incident’) in the second sentence is not only an obligatory constituent of *tæ□bir* (‘interpret’) but can be assumed to be In Focus (in the addressee’s current center of attention) because it was part of the interpretation of the immediately preceding sentence.

Familiar, Activated, and Uniquely Identifiable. The following selection has been selected from the story “Sweet Hat”. Here, two friends have met in the countryside, and are talking about a variety of topics, jumping from one topic to another, and back again.

In the fifth (5th) paragraph into the story, a young woman wearing red trousers is mentioned. Below is the first mention of her, when she is first introduced on stage. In (90) below, the speaker has been looking at a field of trees and flowers, when he spots *yek gol-e atæšin-e bozorg* (‘one big fiery flower’):

- (90)
- | | | | | | |
|--------|--------------|-----|--------------|-----------------|--------------|
| dær | mian-e | yek | sahne-ye | gol-ha-ye | |
| in | middle-LNKPT | one | scene-LNKPT | flower-PL-LNKPT | |
| qermez | ... | yek | gol-e | atæšin-e | bozorg |
| red | ... | one | flower-LNLPT | fiery-LNKPT | big |
| | | | | | did-æm... |
| | | | | | see-PAST-1SG |

‘In the middle of a scene of red flowers ... I saw one big fiery flower...’

In the next paragraph (next sentence), it becomes clear that the “one big fiery flower” is referring to a woman, one who is wearing red pants:

- (91)
- | | | | | | |
|--------------|-------|------------|-------------|-----------------------|-------------------|
| čon | xeili | dur | bud | ne-mi-tævanest-æm | |
| because | very | far | is-PAST-3SG | NEG-DUR-able-PAST-1SG | |
| be-bin-æm | | an | doxtær-e | tonban | qermez, sobh-e |
| SUBJ-see-1SG | | that | girl-LNKPT | trouser | red morning-LNKPT |
| be | an | zudi | dær | mian-e | sæbze |
| to | that | early | in | middle-LNKPT | greenery |
| | | | | | va |
| | | | | | and |
| gol | če | mi-kon-æd. | | | |
| flower | what | DUR-do-3SG | | | |

‘Because it was far, I couldn’t see what that red-trousered girl was doing that early in the morning in the middle of flowers and vegetation.’

Several paragraphs (12 sentences) later, the author comes back to this girl in the story, in (92) below. Here, then, *an doxtær-e tonban qermez* (‘red-trousered girl’) is coded as Familiar, because it was mentioned previously in the discourse; however, the referent cannot be coded higher (Activated or In focus), since it does not meet any of the criterion for higher statuses (e.g., it was not mentioned in previous two sentences—it was mentioned 12 sentences earlier).

- (92) aheste an **doxtar-e** **tonban qermez ra** nešan
 slowly that **girl-LNKPT** **trouser red ra** point
- dad va goft leyla avaz-e ašeqaneh
 give-PAST-3SG and say-PAST-3SG Leyla song-LNKPT love
- mi-xan-æd.
 DUR-sing-3SG
- ‘Slowly [my friend] pointed to that red-trousered girl and said Leyla is singing a love song.’

In the next sentence that follows in the story, in (93), the referent □*u* (‘she’) can be assumed to be In Focus since it refers to Leyla which is in a prominent (subject) position in the previous sentence.

- (93) moqazele-ye gav-ha □**u** **ra** be-yad-e
 love-making-LNKPT cow-PL **3SG** **ra** to-memory-LNKPT
- ešq-æš ændaxt-e [æst].
 love-3SG-POSS throw-PAST-PSPT is-3SG
- ‘The cow’s love-making has reminded her of her own love.’

In the next sentence that follows (93), *in tæsnif* (‘this song’) refers to the song Leyla is singing, two sentences earlier in (92). As such, it can be assumed to be at most Activated:

- (94) **in** **tæsnif** **ra** mæn bælæd-æm
 this **song** **ra** I know-1SG
 ‘I know this song.’

In (95) the referent of the DP *gav-ha-man* (‘our cows’) is coded as Uniquely Identifiable since there is no reason to assume that the referent is already in the addressee’s memory; however, it contains enough descriptive content to pick out not just any ‘cows’, but ‘those that belong to us’.

- (95) emruz mæn væ Morad **gav-ha-man** **ra**
 today I and Morad **cow-PL-Poss-1Pl** **ra**
- ba ham[digær] □æng mi-ændaz-im
 with each[other] war DUR-throw-1Pl
- ‘Today I and Morad will have our cows fight each other.’

4.4 Children's Stories

The set of five children's stories examined yielded 48 *ra*-marked DPs in a data set containing 1,679 words. These children's stories were all published in 1953 in Tehran.

	Word Count	<i>ra</i> Count
'Olaagh va Baar-e Namak' ("Donkey and The Load of Salt")	360	8
'Gorg va Barreh' ("Wolf and Lamb")	380	13
'Khar dar Lebaas-e Shir' ("Donkey in Lion's Clothes")	323	10
'Shekam-e Ghurbaaghe' ("Frog's Stomach")	256	11
'Murcheh va Parvaaneh' ("Ant and Butterfly")	360	6
Totals	1,679	48

Table 18 Number of Words and *ra*-Marked DPs in Children's Stories

Results of examination of the referent of each *ra*-marked DP, according to highest cognitive status, is given in Table 19 below.

	In Focus	Activated	Familiar	Uniquely Identifiable	-i+ra	Not Coded
Donkey and The Load of Salt	4		1	1	2	
Wolf and Lamb	5	2	1	4		1
Donkey in Lion's Clothes	4	3	1	1	1	
Frog's Stomach	1	4	3	3		
Ant and Butterfly			1	4	1	
Totals	14	9	7	13	4	1

Table 19 Highest Cognitive Statuses of *ra*-marked DPs in Children's Stories

Following, one example for each of the cognitive statuses In Focus, Activated, Familiar, and Uniquely Identifiable is offered.

In Focus.

(96)	olaq donkey	ham-in-ke also-this-REL	vared-e enter-LNKPT	ab water	šod, become-PAST-3SG		
	pa-yæš foot-3SG-POSS	pič twist	xord eat-PAST-3SG	væ and	digær thence	na-tævanest NEG-able-PAST-3SG	
	xod self	ra ra	kontorol control	kon-æd. SUBJ-do-PRES-3SG			

‘As soon as the donkey entered the water, his foot twisted and could no longer control himself’

In (96), selected from *Olaagh va Baar-e Namak* (‘Donkey and The Load of Salt’), *xod* (‘self’) refers to the donkey, which is at the center of attention because it was mentioned in a syntactically prominent (subject) position in the preceding clause.

Activated.

The example below, selected from *xær dær lebas-e šir* (‘Donkey in Lion’s Clothes’) provides an example of a DP whose referent is coded as at most Activated.

(97)	heyvan-e animal-LNKPT	bæ□di next	ke REL	češm-æš eye-1SG-POSS	be to	šir lion	oftad fall-3SG-PAST	
	yek one	rubah fox	bud. be-3SG-PAST	dær in	hali state	ke REL	moqabel-e front-LNKPT	šir lion
	xæm bend	šod-e become-3SG-PSPT		bud be-3SG-PAST	væ and	æz from	tærs fear	
	mi-larz-id DUR-shake-3SG-PAST		goft say-3SG-PAST	aqā Mr.	šir lion	šoma you-POLITE	če what	
	yal-e mane-LNKPT	tærsnaki fearsome		darid. have-POLITE	xaheš plead	mi-kon-æm DUR-do-1SG	mæ-ra 1SG-ra	
	næ-xor-id. NEG-eat-3PL	rubah fox	dær in	hali state	in this	hærf word	ra ra	mi-zæd DUR-hit-3SG
	ke REL	æz from	zendegi-e life-LNKPT	xodæš REFLEX		na-omid NEG-hope		
	šode become-3SG-PSPT		bud. be-3SG-PAST					

‘The next animal whose eyes fell upon the lion was a fox. While he had bent down in front of the lion and shaking from fear, said: ‘Mr. Lion, what fearsome mane you have! I

beg you not to eat me.’ The fox was saying **these words**, while he had become hopeless about his life [i.e., he had no hope for continuing to live]’.

The DP *in hærf* (‘these words’—literally, “this word”) refers to *aqa šir šoma če yal-e tærsnaki darid. xaheš mi-kon-æm mæ-ra næ-xor-id*. (‘Mr. Lion, what fearsome mane you have! I beg you not to eat me’). Here, the referent of the DP in question is assumed to be at most Activated, since it is in the addressee’s current short term memory as it was mentioned in the previous sentence. It is argued that referent can be assumed to be at most Activated, and not In Focus, because the DP in question did not occur in a syntactically prominent position (e.g., main clause subject); it was not a higher-level topic in the immediately preceding sentence; and, it was not mentioned in the same sentence.

Familiar.

In (98) below, from the story ‘The Donkey and the Load of Salt’, there is donkey who is carrying salt for a man. Earlier in this story, the donkey has once already dropped the load of salt in the river, resulting in the dissolving of the salt. In doing so, he discovers that his load is lighter and therefore easier to carry. Later on in the story, on a separate occasion, he decides to do the same, hoping to lighten his load once more.

(98)	ab-e water-LNKPT	rud-xane river-house	bar-e time-LNKPT	digær again	næmæk-ha-ye salt-PL-LNKPT
	mærd-e Man-LNKPT	bičare ra poor ra	dær xod in self	hæl dissolve	kærd. do-PAST-3SG

‘The river’s water once again dissolved the poor man’s salt in itself.’

Here, the addressee is expected to remember, and therefore be familiar with, the poor man's salt, since it was mentioned earlier in the story. However, it cannot be assumed to be Activated because it has not been mentioned for some time.

Uniquely Identifiable

In (99) below, while the *æfsar* ('harness') has not been mentioned before, *olaq* ('donkey') had already been activated. As such, it is expected that the addressee is able to construct a unique representation of a harness, by way of a bridging inference to the already activated *olaq* (donkey) which is in the addressee's memory⁵⁸.

(99)	mærd man	æfsar-e harness-LNKPT		olaq donkey	ra ra	ba with	ehtijat care	gereft take-PAST-3SG
	ta so	æz from	rud-xane river-house	ræd pass	šæv-ænd SUBJ-become-3PL			
	'The man took the donkey's harness carefully so [that] they could cross the river.'							

Not Coded: Formulaic Expression.

Lastly, the following DP from 'The Wolf and the Lamb' was not coded for cognitive status, as the phrase *færar ra bær qærar tær ih dad* ('favored flight over fight') is a saying in Persian.

(100)	gorg-e wolf-LNKPT	færib-xord-e deceive-eat-PSTPT		ba with	soræt-i speed-i	ke REL
	mi-tævanest DUR-able-PAST-3SG	færar escape	ra ra	bær over	qærar stationary	tær ih favor
	dad do-PAST-3SG					

'The duped wolf, speedily, favored flight over fight' [i.e., 'ran away as fast as he could']

⁵⁸ It is assumed that the addressee can identify a unique referent (i.e., infer an entity) by linking it indirectly to a recently activated referent.

4.5 Newspaper Articles: ‘Sports’ Section

A total of eight articles from the ‘Sports’ Section of a newspaper called “Iran” were examined (Iran: March 19, 2003). These articles, together, consisted of 2,898 words, consisting of 61 *ra*-marked phrases which were subsequently coded for the highest cognitive status. Table 20 Eight Articles from the Sports Section of 'Iran' Newspaper below outlines the details.

	Word Count	<i>ra</i> Count
‘Nekunaam: Mikhaahim Baa bacheh-haa-ye Paas be Shomaal Beravim’ (“Nekunam: We Want to Go Up North with the <i>Pas</i> Players”)	174	4
‘Payaam-e Peykaan ham Ta’m-e Shekast ra Cheshid’ (“ <i>Payam-e Peykan</i> Also Tasted the Flavor of Defeat”)	82	2
‘Tahavvol dar Fa’aaliyat-haa-ye Varzeshi-e Mantagheh’ (“Change in Sports’ Activities in the Region”)	122	1
‘Persepolis Choob-e Kharid-haa-yash raa khord’ (“ <i>Persepolis</i> Was Punished Due to its Purchases [of new players]”)	950	23
‘Fulaad be Bazikonaanash Eidi Daad’ (“ <i>Fulad</i> ” Gave its Players [New Year] Gifts”)	133	0
‘Aaghaazi Bad, Paayaani Badtar’ (“A Bad Beginning, A Worse Ending”)	943	20
‘500 Hezar Dollar baraa-ye Seh Maah’ (“500 Dollars for Three Months”)	265	7
‘Estili: Faghat Pahlevaan-e Zende h raa Eshgh Ast’ (“Estili: <i>Hurray</i> Only for the Alive Hero”)	229	4
Totals	2,898	61

Table 20 Eight Articles from the Sports Section of 'Iran' Newspaper

Detailed examination of the referent of each *ra*-marked DP is presented in Table 21.

	In Focus	Familiar	Uniquely Identifiable	Not Coded
Nekunam: We Want to Go Up North with the <i>Pas</i> Players	1		3	
<i>Payam-e Peykan</i> Also Tasted the Flavor of Defeat	1			1 (Title)
Change in Sports’ Activities in the Region			1	
<i>Persepolis</i> Was Punished Due to its Purchases	7	4	11	1 (Title)
<i>Fulad</i> Gave its Players Gifts				
A Bad Beginning, A Worse Ending	3	4	12	1 (Formulaic Expression)
500 Dollars for Three Months	2	1	4	
Estili: <i>Hurray</i> Only for the Alive Hero	2		2	
Totals	16	9	33	3

Table 21 Highest Cognitive Statuses of *ra*-Marked DPs in Sports Section

In Focus

In the following selection below, the reflexive pronoun *xod* (‘self’) refers to Hejazi, who was just mentioned.

(101)	čera why	he□azi Hejazi	bajæd must	fæqæt only	xod self	ra ra	mostæhæq-e deserving-LNKPT
		moræbi-gæri ⁵⁹ -e Coach-‘ician’-LNKPT		esteqlal Esteghlal		be-danæd? SUBJ-know-PAST-3SG	

‘Why must Hejazi consider only **himself** [as] deserving of coaching Esteghlal?’

Since ‘Hejazi’ was mentioned earlier in the same clause, it is assumed to be in addressee’s current center of attention, and therefore assumed to be In Focus.

Familiar.

(102) below, from the sports section, demonstrates a situation in which the referent of the DP, ‘Esteghlal’, is assumed to be Familiar because it is shared cultural knowledge. ‘Esteghlal’ is one of the two national soccer teams in Iran. Considering that Iranians are fervent followers of soccer, and this article appears in the sports section of a major newspaper, it is assumed that the addressee is expected to be familiar with the Esteghlal soccer team.

(102)	roland Roland	kox Koch	ke REL	æz from	mordad-mah Mordad-month	vared-e enter-LNKPT	iran Iran
	šod, become-PAST-3SG			ta until	axær-in last-SUPRL	ruz-ha-ye day-PL-LNKPT	sal year
	be in	towr-e manner-LNKPT	□eddi serious	esteqlal Esteghlal	ra ra	hedajat guide	kærd. do-PAST-3SG

‘Roland Koch, who entered Iran from the month of Mordad (July-August), guided **Esteghlal** in a serious manner until the last days of the year.’

⁵⁹ Suffix, connoting the practitioner of a profession, akin to the English “-ician”.

Uniquely Identifiable.

In (103) while *sæbzi væ tæravæt* ('greenness and freshness') has not been mentioned before, the addressee can associate a unique representation with the referent of *sæbzi væ tæravæt*, on the basis of the DP itself. As such, the referent is coded as Uniquely Identifiable.

(103)	amædæn-e coming-LNKPT	bæhar spring	ke REL	sæbzi greenness	væ and	tæravæt freshness	ra ra
	be to	hæmrah together	daræd has-3SG	forsæt-i chance-i	æst be-PRES-3SG	ke REL	
	fareq detached	æz from	hæjahu hubbub	nešæst SUBJ-sit-PAST-3SG	væ and	be at	
	□æmæl-kærd-e action-do-PAST-LNKPT	xod self	dær in	jek one	sal year	be-negær-im. SUBJ-look-3PL	

'The coming of the spring which has [i.e., brings] with it the greenness and freshness, is a chance to sit, detached from hubbub, and look at one's actions in one year.'

Not Coded

The following is an example of the DPs that were not coded for cognitive status. In this particular case, the entire phrase is a formulaic expression, used often as a whole in Persian. Note that this same phrase shows up in one of the children's stories (c.f. (100) above.)

(104)	færar flight	ra ra	bær to	qærar stationary	tær□ih prefer	dad. give-PAST-3sg
-------	-------------------------------	------------------------	-----------	---------------------	------------------	-----------------------

'S/he preferred to run rather than stay' [i.e., chose 'flight over fight']

4.6 Newspaper Articles: ‘World’ Section

A total of nineteen (19) articles from the ‘World’ sections of three news sources were selected at random, yielding 4,343 words and eighty five (84) *ra*-marked DPs. One article from the online news source *Iranian News Agency* (March 24, 2003) yielded two (2) *ra*-marked DPs in a data set consisting of 205 words:

	Word Count	ra Count
‘Jang-e Amrikai: Yek Ma’aareez-e Araaghi az Taghiir-e Barnaame-ye Jang-e Amrika Alaihe Araagh Sokhan Goft’ (“American War: An Iraqi Opponent of The American Change of Plans for War Against Iraq Spoke”)	205	2

Table 22 One Article from 'World' Section of IRNA

Twelve articles were selected at random from the ‘World’ Section of March 19, 2003 *Keyhan* (“Universe”), which yielded 41 *ra*-marked DPs in a data set consisting of 2,140. Details follow in Table 23 below:

	Word Count	<i>ra</i> Coun t
'Towte'e-ye Kharaabkari-e Emrikaa-i-haa dar Kubaa Naakaam Mand' ("The American Plot for Destruction in Cuba Remained Unsuccessful")	123	4
'Fard-e Mosallah-e Yamani 2 Karshenaas-e Emrikaai va Kaanaadai raa be Halaakat Resaand' ("Armed Yemeni Killed two American and Canadian Experts")	101	2
'Tashdid-e Tadaabir-e Aminyati Atraaf-e Markaz-e Siyaasi-Nezaami-e Emrikaa dar Daakhel va Khaarej-e Keshvar' ("Repetition of Security Policy Surrounding the Internal and External American Political-Military Center")	264	2
'E'teraaz-haa-ye Jahaani Alay-he Jang Sheddāt Gereft' ("World Objections to Against War Intensified")	322	4
'Neyoyork Zir-e Kontorol-e Niruhaaye Zedd-e-shoresh Gharaar Gereft' ("New York Was Placed Under the Control of Anti-riot Forces")	192	3
'Saddam va Raamsfeld 18 saal ba'd' ("Saddam and Rumsfeld, 18 Years Later")	113	2
'Emrika: Agar Saddam Ham Beravad Hamleh Mikonim' ("America: We Will Attack, Even if Saddam Leaves")	169	11
'Lomond: Behbud-e Chehreh-ye Amrikaa dar Jahaan Gheyr-e-momken Ast' ("Le Monde: Improving America's Reputation in the World Is Impossible")	121	1
'Dokhtar-e Saalvaador Alandeh Re'is-e Majles-e Nemaayandegaan-e Shili Shod' ("Salvador Allende's Daughter Elected as Speaker of Chile's House")	177	1
'Vakil-e Mohjebah-ye Faraansavi raa be Maraasem-e Tahlif-e Vokalaa Raah Nadaadand' ("They Did Not Allow Covered [with Islamic Hejab] French Lawyer to the Lawyers' Oath Ceremony")	104	3
'Az Tars-e Marg Mordand' ("They Died of Fear of Death")	130	3
'Hemaaghat-e Digar-e Bush' ("Bush's Other Foolishness")	324	5
Totals	2,140	41

Table 23 Twelve Articles from 'World' Section in *Keyhan*

Also, six randomly selected articles from the March 19, 2003 publication of *Hamšæhri* (Fellow Townsman) yielded 1,998 words and 41 *ra*-marked DPs. Details are shown in Table 24 and Table 25 below.

	Word Count	ra Cou nt
'Mozaakeraat-e Iran va Yaman dar baareh-ye Bohraan-e Araagh' ("Discussions of Iran and Yemen about the Iraqi Crisis").	360	8
'Baaztaab-e Jahaani-e oltimaatom-e Bush be Saddam' ("World's Reaction to Bush's Ultimatum to Saddam")	615	18
'Sherkat-haa-ye Nafti-e Jahaan Kharid-e Naft-e Araagh ra Motafaghef Kardand' ("Oil Companies in the World Have Stopped the Purchase of Iraqi Oil")	252	5
'Tasvib-e Ghanoon-e Ekhtiyaar-aat-e Nakhost Vazir-e Felestin' ("Ratification of the Authority Law of the Palestinian Prime Minister")	546	9
'Dar baareh-ye Esteghraar-e Niruhaa-ye Emrikaai Parleman-e Torkiyeh baar-e digar Tasmim Migirad' ("The Turkish Parliament once again Makes a Decision about the Settlement of the American Forces")	152	1
'Kashf-e Selaah va Mavaadd-e Monfajereh dar Riaaz' ("Discovery of Weapons and Explosive Materials in Riyadh")	73	0
Totals	1,998	41

Table 24 Six Articles from 'World' Section in *Hamshehri*

	In Focus	Activated	Familiar	Uniquely Identifiable	-i+ra	Not Coded
American War	1			1		
The American Plot for Destruction in Cuba Remained Unsuccessful	2			2		
Armed Yemeni Killed two American and Canadian Experts				2		
Repetition of Security Policy Surrounding the Internal and External American Political-Military Center			1	1		
World Objections to Against War Intensified			1	3		
New York Was Placed Under the Control of Anti-riot Forces		1	1	1		
Saddam and Rumsfeld, 18 Years Later	1			1		
America: We Will Attack, Even if Saddam Leaves	3	3		5		
Le Monde: Improving America's Reputation in the World Is Impossible	1					
Salvador Allende's Daughter Elected as Speaker of Chile's House	1					
They Did Not Allow Covered [with Islamic Hejab] French Lawyer to the Lawyers' Oath Ceremony	2					1
They Died of Fear of Death	1			2		
Bush's Other Foolishness	1	1	1	2		
Discussions of Iran and Yemen about the Iraqi Crisis	1	1	2	3		1
World's Reaction to Bush's Ultimatum to Saddam	5	4	2	4	1	2
Oil Companies in the World Have Stopped the Purchase of Iraqi Oil	3		1	1		
Ratification of the Authority Law of the Palestinian Prime Minister	7	1		1		
The Turkish Parliament once again Makes a Decision about the Settlement of the American Forces	1					
Discovery of Weapons and Explosive Materials in Riyadh						
Totals	30	11	9	29	1	4

Table 25 Highest Cognitive Statuses of ra-Marked DPs in World Section

In Focus

In (105), *in kešvæ* ('this country') refers to Australia, mentioned earlier in the same sentence. As such it is classified as In Focus because it appears in the subject position in the same clause.

(105)	dowlæt-e government-LNKPT	ostoralia Australia	be to	hæme-ye all-LNKPT		
	diplomat-ha-ye diplomat-PL-LNKPT	□æraqi-ye Iraqi-LNKPT	moqim-e resident-LNKPT	in this	kešvæ country	
	dæstur order	dad give-PAST-3SG	ke REL	be to	sor□æt speed	in this kešvæ ra country ra
	tærk depart	kon-ænd. SUBJ-do-3PL				
	'The Australian government gave orders to all Iraqi diplomats residing in this country to immediately depart this country .'					

Activated.

(106) appears at the very beginning of an article on the world reaction to George Bush's ultimatum to Saddam Hussein. Here □æraq ('Iraq') is mentioned in the first sentence. Two sentences later, □æraq is marked with *ra*. It is argued that the referent is in the addressee's short term memory, and coded as at least Activated because it was mentioned two sentences earlier. Here □æraq is not classified as In Focus because it does not meet any of the criterion for In Focus (i.e., referent not mentioned in a syntactically prominent position in the last two sentences; it was not a higher-level topic of the preceding sentence; and, it was not mentioned earlier in the same sentence.)

(106)	zærb-ol-æ̌æl-e ⁶⁰ stroke-DEF-ultimate-LNKPT	48	48 hour-LNKPT	sǎæte-ye George	̌oř Bush	buš
	ræ̌is president	̌omhur-e republic-LNKPT	amrika America	be to	sæddam Saddam	hosein Hossein
	ræ̌is president	̌omhur-e republic-LNKPT	̌æraq ke Iraq REL	sobh-e morning-LNKPT	færda tomorrow	
	be to	payan end	mi-res-æd, DUR-arrive-3SG	vakoneš-ha-ye reaction-PL-LNKPT	besijar very	motæfavet-i different-i
	dær in	sæth-e level-LNKPT	̌æhan world	bærængixt. provoke-PAST-3SG		
	be in	gozareš-e report-LNKPT	xæbar-gozari-ha, news-place-PL	̌oř George	buš Bush	dær in
	notq-e speech-LNKPT	15 minute-LNKPT	dæqiqei-e minute-LNKPT	xod self	ba with	mottæhem guilty
	sæddam Sadam	hosein Hussein	be with	na-did-e-gereftæn-e NEG-see-PSPT-take-LNKPT		
	qæť-name-ha-ye cut-letter-PL-LNKPT		sazman-e organization-LNKPT	melæl, country-PL		
	towlid-e production-LNKPT		tæslihat-e weapon-PL-LNKPT	koštar-e killing-LNKPT	̌æm̌i, hemajæt	massive support
	æz from	terrorism terrorism	væ and	nefræt hatred	nesbæt regard	be to
	væ and	mottæhed-an-æš allied-PL-3SG-POSS	goft say-PAST-3SG	dæhe-ha decade-PL	zolm tyranny	væ and
	færibkari deceitfulness	inæk now	be to	payan end	resid-e arrive-PSPT	æst be-3SG
	hosein Hussein	væ and	pesær-an-æš son-PL-3SG-POSS	bajæd must	tejje during	48 48
	̌æraq Iraq	ra ra	tærk abandon	kon-ænd do-3PL	væ and	dær in
	xod-dari self-have	æz from	in this	kar, work	bohran-e crisis-LNKPT	̌æraq Iraq
	hæmle-je attack-LNKPT	nezami military		moňærxahæd result have-3SG	šod. become-PAST-3SG	be to

⁶⁰ The construction ‘zærb-ol-æ̌æl’ is an Arabic phrase, used in Persian to mean ‘ultimatum’. ‘ol’ is the spoken version of ‘æl’, the definite determiner in Arabic, restricted in Persian to phrases borrowed wholly.

‘The 48-hour ultimatum of George Bush, America’s President, to Saddam Hussein, Iraq’s President, which ends tomorrow morning, provoked very different reactions on the world level. Based on reports of news organizations, George Bush, in his 15-minute address—charging Saddam Hussein with disregarding U.N. resolutions, production of weapons of mass destruction, supporting terrorism, and hatred towards America and its allies—said [that] decades of tyranny and deceit has now come to an end, and Saddam Hussein and his sons must leave **Iraq** during the [next] 48 hours, and in case of refusal, the Iraqi crisis will result in military attack.’

Familiar and Uniquely Identifiable. The following passage contains two *ra*-marked DPs, the first Familiar and the second, Uniquely Identifiable.

(107)	gerdhard Gerhard	šerowder Schroder	sædr-e top-LNKPT	æ□zæm-e great-LNKPT	alman Germany		
	□æhan ra world ra	astane-ye brink-LNKPT	□æng war	danest know-3SG-PAST	va and	goft said	
	□æng war	□ælæyh-e against-LNKPT	□æraq Iraq	ke REL	bær on	æsær-e result-LNKPT	an that
	hezaran thousands	bi-gonah without-sin	jan-e life-LNKPT	xod self	ra ra	æz from	dæst hand
	xahænd want-3PL-FUTURE	dad give-3SG	tow□ih explanation	pæzir accept	ni-st. NEG-be-3SG		

‘Gerhard Schroder, the Chancellor of Germany, considered [declared] the **world** on the brink of war, and said war against Iraq—which as a result of it thousands of innocents will lose **their life**—is inexplicable.’

The referent of □æhan (‘world’) is classified as Familiar because it is assumed to be shared cultural knowledge between the speaker and the hearer. In the case of *jan-e xod* (‘their life’), the referent was coded as Uniquely Identifiable, since *jan* (‘life’) can be linked to the recently activated *hezaran bi-gonah* (‘thousands of innocents’).

4.7 Newspaper Articles: ‘Culture’ Section

Three articles from the ‘Culture’ section of two different news sources were examined.

Two of the articles came from *IRNA* (Iranian News Agency) and one came from *Aftab-e Yazd* (‘Yazd’s Sunshine’). A total of ten (10) *ra*-marked DP’s were identified from among 880 words. Again, details are provided in the tables below.

	Date Published	Word Count	<i>ra</i> Count
On-line <i>IRNA</i> (Iranian News Agency):	Nov. 5, 2003		
‘Daaneshgaah-haa-ye Keshvar-haa-ye Saheli-e Khazar Tafaahom-naameh-ye Hamkaari Emzaa Kardand’ (‘‘The Universities of the Countries Bordering The Caspian [Sea] Signed a letter of Mutual Understanding of Collaboration’’)		150	0
‘Daryaa-ye Khazar Faghat baraa-ye 100 Saal Markaz-e Tavajjoh va Alaaghe-ye Afkaar-e Omumi Khaahad Bud’ (‘‘The Caspian Sea Will Only Be at the Center of Public Attention and Interest for 100 years’’)		405	3
<i>Aftaab-e Yazd</i> (‘‘Yazd’s Sunshine’’):	Nov. 6, 2003		
‘Abutoraab Khosravi: Djamaalzaadeh Zabaan-e Kukhe va Baazar raa be Adabiyyaat-e Daastaani Tabdil Kard’ (‘‘Abutorab Khosravi: Djamalzaadeh Transformed the Colloquial Language to Literary Fiction’’)		325	7
Totals		880	10

Table 26 Three Articles from ‘Culture’ Sections of IRNA and Aftab-e Yazd

Detailed examination of the referent of each *ra*-marked DP is reflected below.

	Activated	Uniquely Identifiable	Not Coded
‘Daaneshgaah-haa-ye Keshvar-haa-ye Saheli-e Khazar Tafaahom-naameh-ye Hamkaari Emzaa Kardand’ (‘‘The Universities of the Countries Bordering The Caspian [Sea] Signed a letter of Mutual Understanding of Collaboration’’)			
‘Daryaa-ye Khazar Faghat baraa-ye 100 Saal Markaz-e Tavajjoh va Alaaghe-ye Afkaar-e Omumi Khaahad Bud’ (‘‘The Caspian Sea Will Only Be at the Center of Public Attention and Interest for 100 years’’)		3	
‘Abutoraab Khosravi: Djamaalzaadeh Zabaan-e Kukhe va Baazar raa be Adabiyyaat-e Daastaani Tabdil Kard’ (‘‘Abutorab Khosravi: Djamalzaadeh Transformed the Colloquial Language to Literary Fiction’’)	1	4	2
Totals	1	7	2

Table 27 Highest Cognitive Statuses of *ra*-Marked DPs in ‘Culture’ Sections

Below I present a sampling of *ra*-marked DPs from the article titled *Abutoraab Khosravi: Djamaalzaadeh Zabaan-e Kukhe va Baazar raa be Adabiyyaat-e Daastaani Tabdil Kard* ('Abutorab Khosravi: Djamalzadeh Transformed the Colloquial Language to Literary Fiction'). In the title of the article there is a *ra*-marked DP which was not coded as a matter of course. The article begins with the following quote.

(108)	goruh-e group-LNKPT		færhængi: cultural:	“□æmalzade Jamalzade	ævvælin foremost	nevisænde-i writer-i		
	æst is	ke REL	zæban-e language-LNKPT	kuče alley	o and	bazar bazaar	ra ra	be to
	ædæbiyat-e literature-LNKPT		dastani fictional	tæbdil convert	kærd.” did			
	‘Cultural [news] group: “Jamalzade the first writer who transformed the colloquial language into literary fiction.” ’							

While neither the Title, nor this first sentence which was a quotation, were coded for cognitive status (recall that none of the titles and quotations were included in this study), they do impact the analysis of *zæban-e kuče o bazar* ('language of alley and bazaar' [i.e., “colloquial language”] in a subsequent mention of this phrase.

Activated. After the first sentence ((108) above), the article continues as follows. While (109) below is given for context, the *ra*-marked DP is given in (110)

(109)	æbutorab Abutorab		xosrævi, Khosravi		dær in	salm-ærg-e year-death-LNKPT		mohæmmæd Mohammad	
	æli Ali	□æmalzade Jamalzadeh	dær in	goftogu conversation		ba with	xæbærgozari-e news agency-LNKPT		
	fars Fars	ba with	bæyan-e statement-LNKPT		in this	mætlæb subject		goft said	
	jæmalzade Jamalzade		joz□e among	čahar four	nevisænde-ye writer-LNKPT		mohem important		dær in
	□ærse-ye area-LNKPT		ædæbiyat-e literature-LNKPT			dastani fictional		va and	

mohem-tær important-COMPR	æz from	an that	mæbda□e origin
ædæbiyat-e Literature-LNKPT	dastani fictional	dær in	zæban-e language-LNKPT
farsi Persian	æst. is		

‘In an interview with Fars news, in commemoration of the one year anniversary of Mohammad Ali Jamalzade’s death, with the declaration of this utterance⁶¹, Abutorab Khosravi said: Jamalzade is among the four important writers in the area of story literature, and more import than that, he is the originator of literary fiction in Persian.’

Two sentences later, the following occurrence of *zæban-e ædæbiyat* (‘language of literature’) is classified as least Activated, since *zæban-e ædæbiyat* was mentioned two sentences earlier. The occurrence of *zæban-e ædæbiyat* two sentences earlier is sufficient for assuming that the referent is in addressee’s short term memory, justifying the Activated classification. However, since none of the criterion for In Focus are met, we cannot assume the referent is higher than Activated⁶².

(110)	xosrævi Khosravi	ezhar expression	dašt, had	jæmalzade Jamalzade			
	zæban-e language-LNKPT	ædæbiyat literature	ra ra	æz from	an that	fæxamæt privilege	
	birun out	aværd brought					

‘Khosravi said that Jamalzadeh took the **language of literature** out of that privilege [i.e., the domain of the privileged].’

⁶¹ Referring to the utterance at the beginning of the piece, i.e., ‘Jamalzade transformed the colloquial language into literary fiction.’

⁶² The presence of a criterion for a particular classification is sufficient but not necessary for a cognitive status having been met.

Uniquely Identifiable.

In (111), *folkolor-e jame'e-ye iran* ('folklore of the Iranian society') has encoded in it adequate descriptive content for the addressee to construct a new unique representation of what is being referred to.

(111)	jæmalzade Jamalzade	folkolor-e folklore-LNKPT	jame□e-ye society-LNKPT	iran Iran	ra ra	
	dær in	dastan-ha-ye story-PL-LNKPT	xod self	be to	kar work	mi-gereft. DUR-held
	‘Jamalzadeh used the folklore of the Iranian society in his stories.’					

4.8 Speech Data

The data used for this portion of the current study was the audio file of a telephone conversation between two Tehrani speakers. As mentioned in the Methodology section, the audio was recorded by LDC (Linguistic Data Consortium) as part of a project called CallFriend. I transcribed, translated the audio into English, and glossed it for a Prelim paper. The dialogue, consisting of 350 turns (2266 words) between the two interlocutors, included 37 *ra*-marked DP's. Each *ra* -marked DP, as a dependent of the predicator in the VP was coded for Cognitive Status of its referent. The following table details the findings. A selection of examples follows Table 28.

In Focus	Activated	Familiar	Uniquely Identifiable	Referential	-i+ra	Not Coded
8	11	5	5		5	3

Table 28 Highest Cognitive Statuses of *ra*-marked DP's in Speech Data

DPs Not Coded: Formulaic Expressions

In Persian, 'to empty someone's place' is a formulaic expression, used to mean 'to remember someone'. As such, the following occurrence of *ra* in B below was not coded for cognitive status.

- (112) A: hām-in-tor mehmuni-ye. bæle. in-□a
also-this-way party-is yes this-place
- mehmuni-ye o □a-ye šoma-m xeili xali-ye
party-is and place-LNKPT you-also very empty-is

'There are parties, one after another. Yes. Here there are parties, and your place is very empty [i.e., you are missed very much].'

- B: qorban-e šoma. □a-ye ma ro⁶³ xeili
offering-LNKPT you place-LNKPT 1PL ra very
- xali be-kon-in
empty IMP-do-3PL

'Thank you [for missing us]. Keep our spot empty.'

In the following, the phrase 'xoda ra shokr' is a formulaic expression, analogous to 'Thanks be to God' in English.

- (113) A: Færæh □an xub bud?
Farah dear well was
'Was Farah well?'
- B: xub bud ælhæmdolelah xoda ra⁶⁴ šokr
well was thank God God ra thank
'(She) was well, thank God.. Thanks be to God'
- A: xob. xoda ra šokr
good God ra thank
'Good. Thanks be to God'

⁶³ Note that the *ra*-marked DP in (112) would be considered In Focus, and still a "definite," if we were to include it in this study. Again, formulaic expressions were not included as a matter of course.

⁶⁴ It can be argued that the referent of *xoda* ('God') is Familiar in the first DP and at least Activated in the second occurrence.

In Focus:

In “B” below, the referent of *kar* (‘work’) is considered to be In Focus, since it can be assumed to be in the addressee’s current center of attention as it is the subject of A’s question.

- (114) A: tabessun-am baz-æm kar mi-kon-i?
summer-PL-also yet-again work DUR-do-2SG
‘Are you working during the summers again?’
- B: mæn ye mah-e **kar-æm** o
1SG one month-is **work-1SG** **ra**
- šoru kærd-æm
start do-PAST-1SG
- ‘I began **my work** [only] one month ago.’

Familiar, In Focus, and Activated:

In the following exchange, A asks B about a friend’s husband, whom she has not met.

The referent of the first *ra*-marked DP in A (*showhær-e mehræk* (‘Mehrak’s husband’)) is assumed to be Familiar. It is clear from the conversation that the addressee already has a representation of Mehrak’s in memory on the basis of previous expressions. The referent of the second *ra*-marked DP in B is assumed to be In Focus, since it refers to Mehrak’s husband, who has been introduced in A and can be assumed to be at the current center of attention at the time of this utterance. The referent of the third DP in B''' (*liaqæt-eš* (worthy + 3rd Singular)) is assumed to be Activated since it refers to Mehrak who has been mentioned two utterances earlier.

- (115) A: *rasti, šowhær-e čiz-o* did-i?
by the way husband-LNKPT thing/INTRJ [Mehrak]-ra see-2SG
'By the way, did you see Mehrak's husband?'
- B: *mehræk-o⁶⁵?* bæle ...
Mehrak-ra yes
'Mehrak's [husband]? Yes.'
- A': *četor bud?*
how was-3SG
'How was [he]?'
- B': *did-æm. bæd ni[st]... bæd-æk ni[s-t]*
Saw-1sg bad NEG-is-3sg bad-DIM NEG-is-3sg
'I saw [him]. [he] is not bad. [he] is not too bad.'
- A'': *bæd bæče-i⁶⁶ ni-s[t]. mi-g-æn xeyli*
bad kid-i NEG-is-3SG DUR-say-3PL very
bæče-ye Khub-i-e
kid-LNKPT good-i-is
'He's not a bad guy. They say he is a very nice guy.'
- B'': *bæd ni-s[t]. enšalla ke ta axær-esh-æm*
bad NEG-is-3SG God willing REL until end-3SG-also
bæče-ye xub-i ba-š-e
kid-LNKPT good-i SUBJ-be-3SG
'He's not bad. God willing, may he be a good kid till the end.'
- A''': [laughs]
- B''': [laughs] *enšalla ke liaqæt-eš-o*
God willing REL worthy-3SG-ra
dašt-e baš-e
have-3SG-PSPT (SUBJ) be-3SG
'God willing, may it be that he be worthy of her'

⁶⁵ Note that the object head has been dropped here.

⁶⁶ Bache is used here to mean "fellow".

Uniquely Identifiable

In the following utterance, A and B are continuing to talk about Mehrak's husband, when B utters the following.

- (116) □elo-ye mu-ha-š-æm ye xurde kæm-e
 front-LNKPT hair-PL-3SG-also one little less-is
- senn-eš** **o** bištær nešun mi-d-e
 age-3SG/Poss **ra** more show DUR-give-3SG

'[there is] fewer hair in front. Shows **his age** [to be] more [than it actually is]'

In (116) while Mehrak's husband's age has not been mentioned before *senn-eš* ('his [Mehrak's husband's] age') can be assumed to be Uniquely Identifiable via a bridging inference to the recently activated and probably In Focus Mehrak's husband.

4.9 Summary

In this section I summarize my findings, describing the patterns I observed. More specifically, I first describe the Cognitive Statuses most commonly associated with the referents of the nominal expressions studied. Next, I describe and explain DPs co-occurring with both *-i* and *ra*.

Considering the referents of the DP's that occurred with only *ra*, I conclude that the data set examined shows that all (100%) of these referents are Uniquely Identifiable or higher. However, the study has also found that *ra*-marked DPs can be Referential (i.e., indefinite), but only when *ra* co-occurs with *-i*. Further, we observe instances of *i+ra* in DPs with references that were higher than referential. Below, I discuss each of these cases; namely: i) DP's occurring with *ra* only, ii) DP's marked with both *-i* and *ra* that are Referential, and iii) DP's marked with both *-i* and *ra* that are at least Uniquely Identifiable.

4.9.1 The Cognitive Status of *ra*-Marked DPs

The data examined for this study included 17,612 words, 407 (2.31%) of which were the morpheme *ra*. Further, 15 of the DPs marked with *ra* were not coded for cognitive status because they occurred out of context (i.e., in a bulleted item, a title, a quotation, or an adage.) Of the remaining 392 *ra*-marked DPs examined, 119 (30.36%) were In Focus, 57(14.54%) were Activated, 45 (11.48%) were Familiar, 157 (40.05%) were Uniquely Identifiable. Further, 14 (3.57%) occurred with *i+ra*, which will be presented below.⁶⁷ Table 13 repeated for convenience below in Table 29 summarizes the findings, while Table 30 gives percentages associated with the findings in Table 29.

⁶⁷ Note that in calculating percentages, the number of decimal points have been reduced to two places for convenience.

	Word Count	<i>ra</i> Count	Not Coded	In Focus	Activated	Familiar	Uniquely Identifiable	-i+ra
Short stories (including 'Prologue' and 'Introduction')	5,546	167	2	51	25	15	70	4
Children's stories	1,679	48	1	14	9	7	13	4
Newspaper articles: Sports section	2,898	61	3	16		9	33	
Newspaper articles: World section	4,343	84	4	30	11	9	29	1
Newspaper articles: Culture section	880	10	2		1		7	
Speech data	2,266	37	3	8	11	5	5	5
Totals	17,612	407	15	119	57	45	157	14

Table 29 Summary Results of Highest Cognitive Statuses of *ra*-Marked DPs in the Entire Corpus Examined

	Word Count	<i>ra</i> Count	% of <i>ra</i>	Not Coded	<i>ra</i> Coded	IF	% IF	Act	% Act	F	% F	UI	% UI	<i>i+ra</i>	% <i>i+ra</i>
Short stories: Prologue and Introduction	1,036	17	1.64	2	15.00	5	33.33	3	20.00	1	6.67	5	33.33	1	6.67
Short stories: Sweet Hat & Victorious Roman	4,510	150	3.33		150.00	46	30.67	22	14.67	14	9.33	65	43.33	3	2.00
Children's stories	1,679	48	2.86	1	47.00	14	29.79	9	19.15	7	14.89	13	27.66	4	8.51
Newspaper articles: Sports section	2,898	61	2.10	3	58.00	16	27.59		0.00	9	15.52	33	56.90		0.00
Newspaper articles: World section	4,343	84	1.93	4	80.00	30	37.50	11	13.75	9	11.25	29	36.25	1	1.25
Newspaper articles: Culture section	880	10	1.14	2	8.00		0.00	1	12.50		0.00	7	87.50		0.00
Speech data	2,266	37	1.63	3	34.00	8	23.53	11	32.35	5	14.71	5	14.71	5	14.71
Totals	17,612	407	2.31	15	392.00	119	30.36	57	14.54	45	11.48	157	40.05	14	3.57

Table 30 Summary Percentage

4.9.2 DP's with both *-i* and *ra*

DP's marked with *ra* only were all at most Uniquely Identifiable, Familiar, Activated, or In Focus; as such, they are all considered “definite”, since definiteness is regarded as Uniquely Identifiable or higher (c.f., Figure 10). Further, there are 14 DPs that occurred with both *-i* and *ra*. Below I present this data and discuss each instance separately.

The referent of the DPs in the data in (117) - (120) below are considered to be Referential. (117) below is from ‘The Ant and the Butterfly’ demonstrating an example in which *-i* and *ra* co-occur, where it is clear that the speaker expects the addressee to know what *pærvane-ye ziba-i* (‘a beautiful butterfly’) is.

- (117)
- | | | | | |
|-----------|---------|----|--------------|--------------|
| yek-bareh | seda-i | be | guš-æš | xord: |
| one-time | sound-i | to | ear-3SG-POSS | hit-PAST-3SG |
- “aqa murče, sælam. rah ræftæn če sæxt-e!”
Mr. ant hello road go-INF what difficult-be-PRES-3SG
- murče be tæræf-e seda negah kærd,
ant to direction-LNKPT sound look do-PAST-3SG
- pærvane-ye** **ziba-i** **ra** did ke aram
butterfly-LNKPT beautiful-i ra see-PAST-3SG REL slow
- bal mi-zæd.
wing DUR-hit-PAST-3SG
- ‘Suddenly, [the ant] heard a sound: “Mr. Ant, hello. How difficult it is to walk!” The ant looked in the direction of the sound, [and] saw **a beautiful butterfly** who was slowly flapping its wings.’

In the example above, no ‘beautiful butterfly’ has been previously mentioned; however, the addressee is expected to construct a representation of it. Also, subsequent mentions of this butterfly also makes it clear that the speaker has a particular butterfly in mind. In (118) below, it is argued that the phrase *rah-e deraz-i* (‘a long road’) is a single unit marking referentiality. That is, while the “long road” has not been mentioned before, it is

referring to a particular road (i.e., ‘one at the foot of a mountain’), the addressee is expected to construct a representation of such a road, as the road continues to come up as the story continues to develop.

(118)	☐æs-r-e afternoon-LNKPT	yek one	ruz-e day-LNKPT	gærm-e warm-LNKPT	tabestan summer			
	bud. be-PAST-3SG	olaq-i donkey-i	ba with	čænd several	xor☐in saddle	por full	æz from	næmæk salt
	rah-e road-LNKPT	deraz-i long-i	ra ra	dær in	yek one	kuh-paye mountain-foot	dær in	
	piš front	dašt. have-PAST-3SG						

‘It was a warm summer afternoon. A donkey with several saddles full of salt had **a long road**, at the foot of a mountain, ahead of him/her.’

In the selection below from ‘The Donkey in a Lion’s Clothes’, we observe the co-occurrence of *-i* and *ra* again, marking a DP whose referent is considered Referential as well.

(119)	sal-ha year-PL	piš, prior	dær in	yek one	☐ængæl-e jungle-LNKPT	bozorg big	olaq-i donkey-i	zendegi live
	mi-kærd. DUR-do-PAST-3SG			yek one	ruz day	olaq donkey	bæste-i package-i	ra ra
								dær in
	☐adde road	pejda find	kærd. do-PAST-3SG					

‘Years ago, in a big jungle, a donkey lived. One day, the donkey found **a package** on a/the road.’

Once again, while the *bæste* (‘package’) had not been mentioned before, a particular package here is intended, as it continues to come up during the course of the story.

In the selection below from the news articles examined, it is argued that □*æsr-e* □*ædid-i* (‘a new era’) is referring to a particular era, namely that of post-Iraqi crisis, and the addressee is expected, reasonably so, to arrive at this speaker’s intended meaning.

- (120)
- | | | | | | | |
|----------------|-------------------|--------------|------------------|-----------------|------|---------|
| towni | beler | næzdik-tærin | mottæhed-e | □or□ | buš | dær |
| Toni | Blair | close-SUPER | ally-LNKPT | George | Bush | in |
| □ælæse-ye | tow□ihi-e | | parleman-e | | | in |
| meeting-LNKPT | explanatory-LNKPT | | parliament-LNKPT | | | kešvæ |
| | | | | | | country |
| goft | bohran-e | □æraq | □ æsr-e | □ ædid-i | dær | |
| said | crisis-LNKPT | Iraq | era-LNKPT | new-i ra | in | |
| siyasæt-e | beynolmelæli | ræqæm | xahæd | zad. | | |
| politics-LNKPT | international | count | want-3SG-FUT | hit-3SG | | |
- ‘Toni Blair, George Bush’s closest ally in explanatory parliamentary session of this country said [that] the Iraqi crisis will write **a new era** in International politics.’

While the referents of the four DPs above were deemed Referential, the majority of the referents of *i+ra* DPs were Uniquely Identifiable. (121) below, from ‘The Donkey and the Load of Salt’, occurs at the very end of the story, where there has already been ample opportunity for the addressee to have constructed a representation of a particular *bar-e sængin* (‘heavy load’). However, reference here is made to “a heavy load” vs. “the heavy load”:

- (121)
- | | | | | | | |
|-----------------|-----------|------|-------------------|-----------------|------------------|--------|
| væ | be | in | tærtib | olaq-e | tænbæl | ma□bur |
| and | to | this | manner | donkey-LNKPT | lazy | force |
| šod | | | bar-e | sængin-i | ra | dær |
| become-PAST-3SG | | | load-LNKPT | heavy-i | ra | an |
| | | | | | | that |
| rah-e | sæxt | o | tulani | hæml | konæd. | |
| road-LNKPT | difficult | and | long | carry | SUBJ-do-PRES-3SG | |
- ‘And in this manner the lazy donkey was forced to carry a heavy load on that long and difficult road.’

Note that in the above selection, reference is not made to a particular object, but to a whole class or type of thing, namely ‘heavy loads’ in general. In fact, as we will see in

the rest of the data below, there appears to be a common pattern between DPs marked with both *i+ra* and reference to a type of thing in a general sense. The following two examples, (122) and (123), are from Victorious Roman.

(122)	mæn I	væ and	xahær-æm sister-3SG-POSS	æz from	hær any	now □ kind	
	goft-o-gu-i say-and-say-i		ra □ e □ about to	morq hen	væ and	□ u □ e chick	
	ba with	xare □ i-an outsider-PL	mamnu □ prohibited		is-PAST-1SG	bud-im.	
	pak clean	kærdæn do-INFNT	væ and	mo □ ættær aromatize		kærdæn-e to do-LNKPT	
	laneh coop	ba with	mæn I	bud, is-PAST-3SG	dane seed	dadæn give-INFNT	væ and
	□ æmaliat-e Operation-PL-LNKPT LNKPT		pezešk-i physician-LNKPT	væ and	□ ærrah-i-e operation-LNKPT-		
	□ u □ e-ha		ra	xahær-æm		bær □ ohdeh	
	Chick-PL	ra	sister-3SG-POSS		on	responsibility	
	dašt. have-PAST-3SG		hær any	□ u □ e-ye chick-LNKPT	limp-i	šal-i ra	ra
	pa-dar foot-have		mi-kærd DUR-do-PAST-3SG	væ and	bæraye for	hær any	mæræz illness
	dæva-i remedy-i		mi-danest. DUR-know-PAST-3SG				

‘My sister and I were forbidden from any kind of conversation about hens and chicks with foreigners. Cleaning and aromatizing the coop was my responsibility, feeding and veterinary and surgical works of the chicks were my sister’s responsibility. She cured **any limp chick** and she knew a medicine for any illness.

In (122) above, *hær □u□e-ye šal-i* (‘any limp chick’) can be argued to be Uniquely Identifiable; however, what is Uniquely Identifiable is the class of “limp chicks”. In other words, while *□u□e* (‘chick’) in this DP is arguably Activated as it is in the addressee’s short term memory, considering the quantifier ‘any’ (= ‘all’; i.e., ‘all limp chicks’), it can

be argued that the DP here is used in the generic sense. That is, *hær □u□e-ye šal-i* (‘any limp chick’) is used to refer to the whole class of limp chicks. Note that, in this generic sense, the addressee is expected to construct a unique representation of the *type* of referent being described (c.f., Footnote 53). The same argumentation as (122) is made for (123).

(123)	bæraye	in-ke	dæst	æz	šureš	væ	šureš-ian
	for	this-REL	hand	from	riot	and	riot-PL-NOM
	bærdar-i	ʔoz	mæqam-e		qonsul-i	hič	
	take-2SG	aside	position-LNKPT		consul-LNKPT	nothing	
	tohfe	væ	rešve-i	ra	næ-pæzir.		
	gift	and	bribe-i	ra	NEG-IMP-accept-2SG		

‘To give up riots and rioters, accept no gift and bribe other than position of consul.’

Here, *hič tohfe væ rešve-i* (‘no gift and bribe’) is considered Uniquely Identifiable as well. That is, the addressee is expected to associate a unique representation of the *type* of thing being described—that is, *any* ‘gift and bribe’. In (124) below, from Sweet Hat, the addressee is expected to know the general sense of what is being described.

(124)	mæn	in-ruz-ha	æz	sæfær	gorizan-æm	væ	
	I	this-day-PL	from	travel	evasive-1SG	and	
	hič	ʔalæm-i	ra	behtær	æz	konʔ-e	xane
	nothing	world-i	ra	better	from	corner-LNKPT	home
	ne-mi-dan-æm						
	NEG-DUR-know-1SG						

‘These days I am evasive of travel, and know **no world** better than the corner of [the/my] house.’

Once again, *hič ʔalæm-i* (‘no world’) is used to refer to “any world” that the addressee can imagine, and subsequently discard it as an option (i.e., *no world*) for a better place than the corner of the house. Similarly, in (125) we see that while *qesse-i* (‘a story’) is clearly specific, it is not referential, since there could be more than one story involved.

(125)	dær in	ðeld-e volume-LNKPT		ævvæl first	væ and	sevvom third	
	tæqyir-at-e change-PL-LNKPT		ændæk-i small-i		[be] [to]	væsile-ye instrument-LNKPT	
	ostad master	suræt-gereft face-take-PAST-3SG		mæsælæn for example	čænd several	qesse story	æz from
	rəsul Rasul	pərvizi ezafe Parvizi addition	šod become-PAST-3SG		væ and	yek-i one-i	
	do two	nevisənde-ye writer-LNKPT		digær other	hazf, remove	væ and	ya or
	qesse-i story-i	ra ra	ke REL	æz from	nevisənde-i writer-i		
	behtær better	æz from	qesse-ye story-LNKPT		čap publish	šod-e-æš become-PAST-PERFECT-3SG	
	xand-e-bud read-PAST-PERFECT-is-PAST-3SG				væ and	æqide opinion	dašt have-PAST-3SG
	ke REL	bayəd must	be-ðā-ye to-place-LNKPT			qesse-ye story-LNKPT	
	qəbl-i-æš previous-i-3SG		čap publish	šəv-əd become-3SG		be-ðā-ye to-place-LNKPT	
	an that	entexab select	šod. become-PAST-3SG				

‘In the first and third volume, a few changes were made by the master; for example, several stories from Rasul Parvizi were added, and one or two other writers were removed, or else, **a story** that he had read by a better writer than the one he had published, and believed that it must be published instead of the previous story, was selected.’

In (125) above, *qesse-i* (‘a story’) is arguably Uniquely Identifiable, referring to a class of stories that happen to fit a particular description, namely the stories that the “master” had “read by a better writer...”. In this sense, it can be argued that reference is, once again, being made to a type of thing being described.

In (126) - (128), also, we are arguably dealing with DPs referring to things or persons in general. Phrases such as *kæs-i* (‘a person’) and *čiz-i* (‘a thing’) may be used in Persian to refer to something or someone in a somewhat general sense.

- (126) xeyli čiz-a-i -ro ke lazem
 many thing-PL-i-ra REL necessary
- dar-æm æz in-□a mi-xær-æm
 have-1SG from this-place DUR-buy-1SG

‘Many of the things that I need I buy from there.’

- (127) næ, mæn ke ⁶⁸ **hær-či⁶⁹ ra** hesab mi-kon-æm
 no 1SG ‘as for me’ **any-thing ra** calculate DUR-do-1SG
- mi-bin-æm un-□a gerun-tar-e
 DUR-see-1SG that-place expensive-COMP-is

‘No. As for me [or, “for my part”] anything (i.e., goods) I calculate, I see it’s more expensive there...’

Similarly, the DP below in (128) occurs with both *-i* and *ra*, the referent of which is following the pattern previously observed. Here, the speaker is talking about not being able to find someone to take care of her child:

- (128) **kæs-i** **ro** peyda næ-kærd-æm.
 person-i **ra** find NEG-do-PAST-1SG

‘I did not find a person [to take care of my child].’

Here, it is argued that the referent of *kæs-i* (‘a person’) is Uniquely Identifiable, as the hearer can link it to an individual who would be taking care of a child. Again, here *kæs-i* (‘a person’) is being used to make reference to *any* individual who can take care of the child. (128) above occurred in the speech data examined. Here the two interlocutors were talking about one of the participant’s only child. Later on in the conversation, the other participant brings up the subject of whether the other person is considering having a second child.

⁶⁸ ‘ke’ is a Relative marker, but here, it is used idiomatically to mean “for my part”, or “as for me”.

⁶⁹ ‘hær čiz-i’ is contracted in speech to ‘hær či’.

- (129) A: **dovvom-i** **ra** ne-mi-xa-i be-za-i?
second-i **ra** NEG-DUR-want-2SG SUBJ-birth-2SG
 ‘Do you not want to give birth to a **second** [child]?’
- B: næ xanum jun
 NEG Ms. dear
 ‘No, dear’
- A': rast mi-gi?
 truth DUR-say
 ‘Are you serious?’
- B': valla
 ‘yes’
- A'': go-ft-æm šayæd **dovvom-i-r-æm**
 say-PAST-1SG perhaps **second-i-ra-also**
- be-xa-i be-za-i, ye dæfe
 SUBJ-want-2SG SUBJ-birth-2SG one time
- do** **ta-i -ro** ba hæm bozorg kon-i
two **CL-i-ra** with also raise do-2SG
- ‘I thought that perhaps you might want to give birth to a **second** one also, raising both [children] at the same time.’

In the exchange in (129) the two interlocutors are talking about B’s only child, when A brings up the subject of whether B is planning on having a second child. While there is no previous mention of a second child, in the use of *dovvom-i* (‘a second’) the hearer can be assumed to know the type of thing being described, i.e., a second child. Similarly, the use of *do ta-i* (‘two-CL-i’) is to signal reference to ‘two children’ in the general sense; that is, the type of thing (i.e., child) being described.

While the data on *i+ra* is limited here, it can be tempting to argue that perhaps we are dealing with a situation in which the referents of these particular DPs are either referential, making reference to particular object, or else they may be generic, collapsing Uniquely Identifiable and Type Identifiable, signaling a unique representation of a type of thing being described. However, such conclusion requires a separate study, outside the scope of this current work. While the current study concentrated on *ra*-only marked DPs,

a future study would have to first consider the role of the referential $-i$, prior to considering how it may interact with ra . As you may recall, the role of $-i$ itself remains unclear. I will come back to the functions of $-i$ and the potential for a future study in Section 5.2 below.

In this section I concluded that the referents of all ra -only marked DPs found in the data studied were Uniquely Identifiable or higher. I also addressed all $i+ra$ DPs individually. While no definitive conclusion was made regarding these DPs, a future study is suggested. The last Chapter of this dissertation, next, is dedicated to overall conclusions.

5 Conclusion

This chapter summarizes the main findings of the dissertation, describing the results of this study, and explaining how the results relate to expectations and to literature cited in the introduction. I also describe what additional research might be in order, and suggest how the results of this study fit into a broader context.

5.1 Summary

This thesis set out to investigate the referential status of DPs marked with *ra*. The reason for undertaking this study was to further understand the nature of this morpheme, as there is a debate surrounding the function(s) of *ra* in literature. While the focus of this study was the referential status of *ra*-marked DPs, as background information, it was noted that there are a considerable number of different varieties of Persian, displaying varying degrees of divergence from the standard variety. To that end, the data I collected and analyzed were drawn from a variety of sources and genres to reflect this diversity. In considering the referential aspect of *ra*-marked DPs, I employed the scalar Givenness Hierarchy framework. It was shown that this framework makes it possible to determine the cognitive (memory and attention) status of the referent assumed of the addressee at a given point in discourse, using concepts more primitive than that of *specificity* and *definiteness* (Gundel, Hedberg, Zacharski 1993).

The data examined in this study showed that the referent of DPs marked with *ra* are *uniquely identifiable* or higher (i.e., they are definites). This conclusion supports claims that *ra* is a marker of definiteness and not merely specificity. See Table 31 below,

repeated here for convenience. As indicated earlier, for a referent to be specific, it only needs to be either *referential* or higher in the Givenness Hierarchy, whereas the more restricted *definite* referents need to be *uniquely identifiable* or higher.

	In Focus	Activated	Familiar	Uniquely Identifiable
Totals	119	57	45	157

Table 31 Summary of Highest Cognitive Statuses marked by *ra* in the Data

These results are consistent with Sadrai (2003; 2008). The results also corroborate the explication of *ra* as reflected in Phillott (1919), Lambton (1953), Ghomeshi (1997), and Mahootian (1997). They do not support the claim that *ra* only marks specificity (Browne 1970, Windfuhr 1990, Karimi 1990). Further, the results in Table 31 show that the vast majority of *ra*-marked DPs are at the two ends of the scale, i.e. at most Uniquely Identifiable or In Focus continuum. As noted earlier, Mahootian (1997) “consider[s] object noun phrases on a scale of most definite to least definite, where *ra* marks object NPs [sic] toward the higher, more definite end of the scale.” Also recall that Mahootian does not offer a framework to determine a “scale” in support of the claim. Considering the referents of *ra*-marked DPs using the Givenness Hierarchy, we can observe that the results of this study, as reflected in Table 31, show that while there is a large number (119) of DP’s whose referents are In Focus, there are also 157 DP’s whose referents were at most Uniquely Identifiable, i.e. not Familiar or Activated. Given this, it appears that the vast majority of *ra*-marked DPs do not “occur on the more definite end of the scale”, even if *familiar* and *activated* referents are in some sense more definite than ones that are at most *uniquely identifiable*. Further investigations are needed to determine this conclusion. At this point, I will suffice it to claim that the majority of the referents of the

ra-marked DPs in this study occurred at the two extreme ends within the “definiteness” scale: In Focus and Uniquely Identifiable.

5.2 Future Research

In this section, I will address several questions outlining the future work to be done on the information status of DPs marked with *ra*: the referential status of *-i*; the referential status of DPs occurring with both *-i* and *ra*; and, the role of dialectal differences and the referential status of *ra*-marked DPs.

5.2.1 The Referential Status of DPs Marked with both *-i* and *ra*

As indicated in Section 4.9.2, there were 14 DPs that occurred with both *-i* and *ra*. The limited data showed that the DPs co-occurring with these two morphemes were at most referential. This finding potentially poses a challenge to the finding that *ra* marks DPs that are at least uniquely identifiable, i.e. that it marks definiteness. However, it could be argued that in the 'i-ra' - marked DPs, a "definite"(at least Uniquely Identifiable) is nested inside an "indefinite" (at most Referential) . The fact that the cognitive statuses 'uniquely identifiable ' and 'referential' are implicationally related, the former entailing the latter, but not vice-versa, allows for this type of a relationship, whereas the “definite”/“indefinite” model suggests discrete, mutually exclusive categories. This proposal will be further developed in future studies of the 'i-ra' marked DPs.

While the DP's with both *i* and *ra* were each presented separately in Section 4.9.2, in that section it was concluded that further research is needed for any definitive conclusions to be drawn. Such undertaking would require that the referential status of *-i* be determined first, before considering the cognitive status of referents of *i+ra* DPs. Traditionally, *-i* has been claimed to mark indefiniteness. As we have seen *-i* can co-occur with *ra*. This co-

occurrence, as pointed out, can potentially pose a problem for researchers who work within the definiteness/specificity framework. The question is, how can a definite marker (i.e., *ra*) co-occur with an indefinite marker (*-i*). The status of *-i*, however, remains disputed. Windfuhr (1990) argues that *-i* marks specific or non-specific indefiniteness: “it marks restrictive selection out of a generic unit or out of a plurality”. The morpheme *-i* can co-occur with both singular and plural nouns. He offers the following examples (with no context)⁷⁰:

- (130) ketab-i
 book-IND
 ‘some/a book’
- (131) ketab-ha-i
 book-Pl-IND
 ‘some books’
- (132) ab-jow-i
 water-barley-IND
 ‘some/a beer’
- (133) ab-jow-ha-i
 water-barley-IND
 ‘some kinds of beer’

Windfuhr states that the fact that *-i* selects a subset out of genericity becomes clear when looking at compound verbs, and offers the following:

- (134) kar mikonæm
 work DUR-do-1Sg
 ‘I am working’
- vs.
- (135) kar-i mikonæm
 work-IND DUR-do-1SG
 ‘I am doing something/some work, I am working some/a little’

⁷⁰ Note that Windfuhr does not provide glosses, but I have included them here for clarity

Windfuhr offers the following example (Numbering is mine) to show that *-i* can be specific or non-specific:

- (136) dombal-e aparteman-i migærdæm
 after apartment-i PRES-look-1sg
 ‘I am looking for an apartment’

He notes that (136) can either imply a specific apartment (e.g., the one I read about in the papers), or any apartment (non-specific). Since definiteness is not necessarily marked in Persian, a noun without *-i* can be interpreted as generic or as definite specific. Windfuhr claims that “the presence of *-i* eliminates genericity,” and states that if we remove *-i* from (136) as in (137) below, we get genericity⁷¹.

- (137) dombal-e aparteman migærdæm
 after-e apartment PRES-look-1sg
 ‘I am apartment-hunting’

While Windfuhr claims that the morpheme *-i* marks indefiniteness, specific or non-specific, Mahootian (1997) claims that one of the ways in which NPs can be marked for indefiniteness is the use of *i* “which indicates [- definite] and [+ specific]”.

Mahootian also claims that the “[m]arking of indefiniteness [whether through the use of *i*, or *ye(k)* ‘one’, or the co-occurrence of *ye(k)* and *-i*] is obligatory for both subject and object noun phrases. Unmarked object noun phrases are interpreted as definite or

⁷¹ This is not a common use of the term genericity. (137) does not mean ‘I am looking for all apartments. Also, it’s not clear how it is different from non-specific interpretation. Generic is a semantic property of the referent (referring to a whole class as opposed to specific members of the class) independent of hearer knowledge. Genericity and definiteness are therefore not viewed as mutually exclusive by all researchers.

generic.” Mahootian does not say anything about bare subject DPs. She offers the following in support (Mahootian’s 328-330)⁷².

- | | | |
|-------|---|-------------------|
| (138) | ye durbin
one camera
‘a camera’ | [Mahootian’s 328] |
| (139) | durbin-i
camera-IND
‘a (certain) camera’ | [Mahootian’s 329] |
| (140) | ye durbin-i
one camera-IND
‘a (certain) camera’ | [Mahootian’s 330] |

Once again, without having a context for these examples, it is difficult to judge the exact nature of the function of these morphemes, and possible interpretations (as the reader may note, the translation provided by Mahootian for (139) and (140) are exactly the same.)

It is also noteworthy that ra is compatible with all three examples given above:

- (141) **ye durbin-o** avordæm [væli ‘un yeki-o næivordæm]
 ‘I brought one camera [but not the other one]’
- (142) **durbin-i-o** xæridæm [ke diruz ba hæm ræftim didim]
 ‘I bought the camera [that we saw together yesterday]’
- (143) **ye durbin-i-o** xæridæm [ke betunæm deraz moddæt æzæš estefadeh konæm]
 ‘I bought a camera [that I could use for the long-run]’

As noted above, some researchers have claimed that indefiniteness in Persian can be marked by the morpheme *-i*. However, *-i* can co-occur with *ra*, as illustrated below:

⁷² It is not clear if Mahootian's claim is that *-i* only marks specific indefinites since the example given shows that it can mark specific indefinites, but Mahootian does not show that it could not mark non-specific indefinites.

- (144) ketab *-i* *ra* (ke diruz xærîde bud) be-hem dad
 book *-i* *ra* (that yesterday bought was) to-me gave-3sg
 's/he gave a/some (specific) book to me' (e.g., that she'd bought yesterday)

As pointed out earlier, this poses a problem for researchers in that there is no way to reconcile the co-occurrence of these two morphemes if one was considered a 'definite' marker (*ra*) and the other an 'indefinite' marker (*-i*) since these concepts are considered mutually exclusive. The question is, how could a definite marker co-occur with an indefinite marker. Windfuhr (1990) resolves this puzzle by claiming that "*ra* is compatible with the indefinite marker *i*, if the latter is specific and implies a unique referent 'a certain, some'."

The above discussion is to simply point out that there are unresolved issues surrounding the morpheme *-i*. Again, although *-i* is not investigated in the current study, it is important to point out that here, too, there may be terminological issues, stemming from lack of precise definitions and not providing a context for examples given. It is difficult to determine what is meant by specificity, definiteness, and genericity without clear definitions and without the context in which given examples may occur. A future study may examine *-i* in naturally occurring discourse, using the more precise Givenness Hierarchy framework. Furthermore, since the cognitive statuses are implicational by definition, the Givenness Hierarchy (GH) will potentially allow us to resolve the apparent contradiction arising from the application of dichotomous, discrete frameworks traditionally used to explicate the referential functions of *ra*. I believe without this type of investigation on *-i* it may be difficult to investigate nominal expressions marked with *i+ra* and *yek* ('one') + *i* + *ra*.

5.2.2 Definite DPs not marked with *ra*

Another relevant question in studying the referential status of *ra*-marked DPs is whether *all* complements (i.e., ‘direct objects’) whose referents are Uniquely Identifiable or higher are marked with this morpheme. Though the question remains to be further investigated in future studies, here is one example that stood out in the speech data.

- (145) A: emšæb ræft-æm **mahi** xærid-æm...
 tonight go-PAST-1SG **fish** buy-PAST-1SG
 ‘Tonight I went [and] bought fish.’
- B: ahan
 ‘Ahh!’ [“I see”]
- A': dašt-æm **mahi** pak mi-kærd-æm
 have-PAST-1SG fish clean DUR-do-PAST-1sg
 ‘I was cleaning [the?] fish’

The referent of *mahi* (‘fish’) in A is at most referential or type identifiable, and *ra* would therefore be judged ungrammatical in the context (i.e., the hearer might say “what fish are you talking about?”)

- (146) # emšæb ræftæm mahi **ra** xæridæm

However, *mahi* (‘fish’) in A', where it apparently refers to the fish that was mentioned or can be inferred from A and it is therefore Uniquely Identifiable is grammatical both without *ra*, as in (145), or with *ra*, with the same interpretation, as in (147)

- (147) daštæm mahi **ra** pak mikærdæm

If the claim is that all definites, i.e., Uniquely Identifiable or higher, must have *ra*, then it appears that we have one counterexample here, as ‘mahi’ in A2 is definite, i.e. at least uniquely identifiable (in fact activated) here, yet it does not require *ra*. And, since all definites are specific (uniquely identifiable implies referential) we also have a

counterexample to the weaker claim that all specific DOs must be *ra*-marked. How then do we explain this? A possible explanation may go something like this: *mahi* in the second utterance is not intended to refer to the same fish as mentioned in the first citation. It is meant in the general sense, i.e., ‘I was engaged in fish-cleaning’. What is important here is the fact that she was busy with something (i.e., cleaning fish) and that she was not going to bed anyway; the information about whether it was the same fish that she just mentioned she had bought was not important in this discourse.

Further examining the rest of the *ra*-marked DPs in the speech data, to see whether *ra* was obligatory each time it occurred, I determined that they were indeed all required, even in cases in which one could argue that *ra* was co-occurring with a phrase that could arguably be in an adjunct position (See Appendix C). Of course, not finding *ra*-marked DPs in naturally occurring data that could retain their grammaticality once *ra* is removed does not mean that *ra* is always obligatory. It only means there were no examples of optional *ra* in the speech data. The next step is to consider this question in the written data used for this dissertation.

5.3 The Bigger Picture

This last section is intended to consider the findings in this dissertation in a broader context by considering theoretical implications and practical applications. The research on *ra* is important in several ways. Aside from implications in teaching and learning Persian as a second/foreign language, an explicit explanation of functions of *ra* has practical applications in computational linguistics. For example, current automatic translation systems, at best, are lacking when it comes to information structure and reference resolution. Further, the results of this study can be used as the starting point for considering the referential status of other morphemes in Persian, many of which still remain enigmatic.

In conducting this study, there remains one last subject I would like to discuss in the context of potential future studies, born out of the process of studying *ra*-marked DPs in this dissertation: the role of dialectal differences that exist today throughout the Persian speaking communities in Iran. A preliminary consideration of the data suggests that in future studies, spoken varieties of Persian need to be studied separately from written (standard) varieties.

Tehrani Speech

There appear to be cases in which *ra* (or its variants *-o* and *-ro*.) do not occur where they are required. To my knowledge, this phenomenon has never been reported in literature. There were three instances of this situation. Note that for the purposes of this study, these DPs were not considered, as they were not realized on the surface level (i.e., all instances

in which *ra* actually occurred in DP were identified and analyzed). In the data presented below, *ra* was not present, where one would expect it.

- (148) æmma mehrak xeili sær-e. be xosus ælan
but Mehrak very superior[looking]-is to regard now
- xod-eš-æm** (i.e., contracted *xod-e-š ra hæm*) laqær kærð-e
self-3SG-ra?-also skinny do-PSPT
- ‘But Mehrak is much more superior [i.e., better-looking]; especially now that she has made herself skinnier [i.e., lost weight]’
- (149) ettefaqæn un-□a čændin bar
coincidentally that-place several times
- hæm-digær-æm** (or alternatively: *hæm-digæ-r-æm*) dide bud-æn..
each-other-ra-also see-PSPT was-3PL
- (150) šæmsi **xanum-æm** dide bud-æn.
Shamsi Ms.-also see-PSPT was-3PL
- ‘It so happens that they had seen each other there several times. Ms. Shamsi had seen them also’ (or, “They had seen Ms. Shamsi as well.”)
- (151) šæmsi **xanum-ø**⁷³-æm dide bud-æn.
Shamsi Ms. -ra -also see was-3PL
- ‘They had seen Ms. Shamsi as well’⁷⁴

Considering careful speech and native speakers judgements, “**xod-eš-æm**” in (148) may be considered a contracted from *xod-e-š ra hæm*. Similarly, in careful speech the ‘bolded’ DP in (149) would have been *hæmdigær ra hæm* ([r] [ra] [h] → [r]). It is not entirely clear if the phonologically realized surface [r] in ‘hæmdigæræm’ is the one that goes with *hæmdigær* or the one the goes with *ra*.

⁷³ **ra** here, too, is contracted.

⁷⁴ Note that this sentence in isolation is ambiguous in that it can be interpreted as ‘Ms. Shamsi had seen them’; however, the context dictates that ‘Ms. Shamsi’ here is in direct object position. Indeed, ‘Ms. Shamsi was in object position, there would be no need for **ra**. However, native judgment tells us otherwise.

In (150) the DP ‘šæmsi xanum-æm’ can be interpreted in two ways:

- (i) šæmsi xanum-æm
(ii) šæmsi xanum-ø-æm (i.e. Careful: šæmsi xanum-ra-hæm)

In case (i), ‘šæmsi xanum-æm’ (careful speech: ‘šæmsi xanum hæm’), would have to be interpreted as the subject and the agent of the sentence ((150) is repeated below, for convenience):

- (152) šæmsi xanum-æm dide bud-æn.
Shamsi Ms. -also see-PSPT was-3PL
 ‘Ms. Shamsi had also seen them’

Whereas is (ii) (careful speech: *šæmsi xanum ra hæm*), ‘šæmsi xanum’ is the object of the sentence, as in (153).

- [illegible]

In the context of the utterance in the speech data, (ii) above is the more plausible interpretation and, as the referent of ‘šæmsi xanum’ is Activated, *ra* is therefore obligatory.

Further investigation is necessary to shed light on this phenomenon. For now, relying on my own native judgment and that of my informants, I claim that in both examples above, *ra* is obligatory in careful speech. As to whether *ra* is phonologically reduced or contracted, further investigations are necessary. An adequate description of the functions of *ra* should help with these interesting cases in examples presented above. For example, a native speaker would judge ***xod-eš-æm** in DO position as (syntactically)

ungrammatical without *ra*. Evidence for the claim that it is contracted must be gathered and presented in subsequent studies of this topic.

Observing examples such as the ones we see above, a relevant question here is why *ra* is not occurring in a position where it is expected to do so. Note that the differences between careful speech and Tehrani examples we have seen here are not limited only to Tehrani dialect. There are a number of other dialects that appear to exhibit similar behaviors. Consider, for example, the formal sentence:

- (154) **dozdi** **ra** hæm be □u nesbæt dad-ænd
 thievery **ra** alao to s/he relation give-3PL
 ‘They connected him/her to [i.e., accused her/him of] robbery as well.’

The Golpayegani version of the utterance in (154) is the following:

- (155) dozzi-a:m be-šeš veslund-æn
 thievery-ra?-also to-3SG connect-3PL
 ‘They connected him/her to [i.e., accused her/him of] robbery as well.’

Several years ago, Professor Iraj Bashiri provided me with about a dozen sentences from an Iranian T.V. series called *særbedaran* (Heads on the Gallows) in Persian, in which one of the characters, while speaking with perfect Persian accent, used sentences that did not contain *ra* at all, where speakers of Standard Persian would clearly say they were required. At the time, I did not know what to think of this. The only explanation I had for this phenomenon was that the script written for this character in the film somehow wanted to signal this individual linguistically different from the others. To what purpose, I did not know. While I still do not have an explanation for what the reason was for having a character that drops his *ra*’s, since then I have come across a very interesting

claim about a dialect in the Kerman province in the south eastern part of Iran. Dehghani (1998) in a survey of Jorft and Kahnoot dialect reports that this dialect does not have *ra* and there are no other morphemes in its stead:

æz an□a ke zabanha va guyešha besuye sadegi piš mirævænd væ dær in rah
 čizhai ra hæzf væ čizhai ra ezafe mikonænd, dær in guyeš “ra” □ælamæt-e
 mæf□ul æslæn vo□ud nædaræd væ hič □ælamæte digæri hæm □aygozine an
 nemišævænd.

‘Considering that languages and dialects move towards simplicity and in this way they omit some things and add some things, in this [Jiroft and Kahnoot] dialect the object marker ‘ra’ does not exist at all, and no other marker replaces it.’

Given what I have observed in the contraction of *ra* in Tehrani and Golpayegani, for example, and in light of Dehghani’s report, a comprehensive study of Persian must necessarily take into account dialectal differences and “native” grammatical judgments. While the study of written/formal varieties is interesting, most linguists are interested in spoken language. As noted in the background section, the written/formal variety used in Persian speaking communities in Iran has changed very little in the past 1000 years, compared to spoken varieties. Future studies on *ra* may have to be limited to specific dialects. For example, written data (newspaper articles, books, etc.) will have to be studied separately from, say, the Tehrani dialect, and so on. Given the morpho-syntactic differences between local dialects compared to the standard varieties, it is plausible that there may be pragmatic (referentiality) differences as well. Future studies on *ra* will have to take this phenomenon into account.

In this study I considered a number of different sources for my data which included spoken Tehrani, as well as various genres in the written/formal varieties. I noted that the Tehrani dialect was the closest to that of what might be considered the “standard” variety,

found in written/formal discourse. I found that the referent of all DPs marked with only *ra* (without *-i*) were Uniquely Identifiable or higher (“definite specific”). I also pointed out that the situation is more complex in DPs that contain *i+ra*, and that the status of *-i* needs to be further investigated before any conclusions about *i+ra* DPs can be proposed. Further, I noted some differences between the Tehrani spoken variety and the standard variety, and presented evidence that several other dialects function differently as well, including one that does not make use of *ra* (or any other morpheme in its stead) at all. I now look forward to future work, taking off where I left off in this current work.

Bibliography

- Allwood, Jens, Lars-Gunnar Andersson, and Östen Dahl (1979). *Logic in Linguistics*. Cambridge: Cambridge University Press.
- Babak, Ali (1996). *Bærræsi-e Zæban Šenazti-e Guyeš-e Zærænd* ("Linguistic Study of Zarand Dialect"). Kerman (Iran): Center for Kerman Studies.
- Bashiri, Iraj (1991). *Persian for Beginners*. Minneapolis: Sorayya Publishers, Inc.
- Browne, W. (1970). More on definiteness markers: Interrogatives in Persian. *Linguistic Inquiry*, Volume 1, No. 3.
- Carnie, Andrew (2002). *Syntax: A Generative Introduction*. Great Britain: U.S., U.K.: Blackwell Publishers.
- Commons Wikimedia, http://commons.wikimedia.org/wiki/File:Iranian_languages_area.png (3/11/2010)
- Crystal, David (1997). *Dictionary of Linguistics and Phonetics*. Fourth Edition. Great Britain: Blackwell Publishers.
- Dehghani, Eslam-e Niknafs (1998). *Bærræsi-e Guyeš-e ĩiroft væ Kæhnuĩ* ("Survey of Jiroft and Kahnooji Dialect"). Kerman (Iran): Center for Kerman Studies
- Ghomeshi, Jila (1997). *Topics in Persian VPs*. *Lingua*, Volume 102, Elsevier.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. (1993). *Cognitive Status and the Form of Referring Expressions in Discourse*. *Language*, Volume 69, Number 2.
- Karimi, Simin (1989). *Aspects of Persian Syntax, Specificity, and the Theory of Grammar*. University of Washington doctoral dissertation.
- Karimi, Simin (1990). *Obliqueness, Specificity, and Discourse functions: Ra in Persian*. *Linguistic Analysis*, Volume 20, Number 3-4.

- Lambton, Ann K. S. (1953). *Persian Grammar*. Cambridge: The Syndics of the Cambridge University Press.
- Lazard, Gilbert (1975). *The Rise of the New Persian Language*. In Frye, R. N. (ED.), The Cambridge History of Iran, V. 4: The Period from the Arab Invasion to the Saljuqs (pp. 595-633). Great Britain: Cambridge University Press.
- Mahootian, Shahrzad (1997). *Persian*. London and New York: Routledge.
- Menashri, David (1992). *Education and the Making of Modern Iran*. Ithaca and London: Cornell University Press.
- NationMaster.com. 08/04/2010). <http://www.statemaster.com/encyclopedia/Image:Persian-Language-Location-Map1.png>
- Peterson, D. J. (1974). *Noun Phrase Specificity*. University of Michigan doctoral dissertation.
- Partee, Barbara H., Alice ter Meulen, and Robert E. Wall. (1987). *Mathematical Methods in Linguistics. Studies in Linguistics and Philosophy*. Netherlands: Kluwer Academic Publishers.
- Phillott, D. C. (1919). *Higher Persian Grammar*. Calcutta: University of Calcutta, published at the Baptist Mission Press.
- Razmjoo, H., M. M. Rokni, et al. (1992). "Farsi." 212.
- Sadrai, Mahmoud (1998). *ra: A Lexical Item in the Context of Utterances*. University of Minnesota Masters Thesis.
- Sadrai, Mahmoud (1999). *Predicate Analysis of Persian DPs: The Case of ra and -i*. University of Minnesota Semantics course paper.
- Sadrai, Mahmoud (2003). *A Cognitive-Based Analysis of ra-Marked DPs in Modern Persian Discourse*. University of Minnesota plan B paper.
- Sadrai, Mahmoud (2008). *Cognitive Status of ra and the Givenness Hierarchy*. University of Minnesota Written Preliminary Exam.

Sadrai, Mahmoud (2009). Codeswitching at the ra-marked Kase Phrase (KP) boundary. University of Minnesota Written Preliminary Exam.

Saffar Moqaddam, Ahmad (2008). Persian Language. 4 vols. Tehran, Iran: Showraaye Gostareshe Zabaan va Adabiyaate Farsi.

Shokouhi, H., and P. Kipka (2003). *A discourse study of Persian ra*. Lingua, 113 (pp. 953-966).

University of Texas at Austin, Library website (4/7/2010).

http://www.lib.utexas.edu/maps/middle_east_and_asia/iran_peoples_82.jpg

Windfuhr, Gernot L. (1979). *Persian Grammar. History and State of its Study*. The Hague, Paris, New York: Mouton Publishers.

Windfuhr, Gernot L. (1990). *Persian*. In Comrie, B (ED.), The world's Major Languages (pp.523-546). New York: Oxford University Press.

Windfuhr, Gernot L. (2010). *The Iranian Languages*. Routledge: London & New York.

Appendices

Appendix A

Source	Date Published	Number of Words	Number of <i>ra</i> Occurred	IF	Act	Fam	UI	<i>i+ra</i>	Not Coded
'Sea of Jewels' ('Daryaa-ye Gowhar')	"Winter", 1988								
Prologue		460	6	2	1	1		1	1
Introduction		576	11	3	2		5		1
Short Story: 'Sweet Hat' ("Shirin Kolaa").		2505	89	30	17	9	32	1	
Short Story: 'Victorious Roman' ("Fateh-e Rumi").		2005	61	16	5	5	33	2	
Totals		5546	167	51	25	15	70	4	2
Newspaper Articles									
Iran Newspaper: Sports Section	19-Mar-03								
'Nekunaam: Mikhaahim Baa bacheh-haa-ye Paas be Shomaal Beravim' ("Nekunam: We Want to Go Up North with the Pas Players")		174	4	1			3		
'Payaam-e Peykaan ham Ta'm-e Shekast ra Cheshid' ("Payam-e Peykan Also Tasted the Flavor of Defeat")		82	2	1					1
'Tahavvol dar Fa'aaliyat-haa-ye Varzeshi-e Mantagheh' ("Change in Sports' Activities in the Region")		122	1				1		
'Perspolis Choob-e Kharid-haa-yash raa khord' ("Persepolis Was Punished Due to its Purchases [of new players]")		950	23	7		4	11		1
'Fulaad be Bazikonaanash Eidi Daad' ("Fulad" Gave its Players [New Year] Gifts")		133	0						

'Aaghaazi Bad, Paayaani Badtar' ("A Bad Beginning, A Worse Ending")		943	20	3		4	12		1
'500 Hezar Dollar baraa-ye Seh Maah' ("500 Dollars for Three Months")		265	7	2		1	4		
'Estili: Faghat Pahlevaan-e Zنده raa Eshgh Ast' ("Estili: Hurray Only for the Alive Hero")		229	4	2			2		
Totals		2898	61	16	0	9	33	0	3
Newspaper Articles: Cultural Section									
On-line IRNA (Iranian News Agency)									
'Daaneshgaah-haa-ye Keshvar-haa-ye Saheli-e Khazar Tafaahom-naameh-ye Hamkaari Emzaa Kardand' ("The Universities of the Countries Bordering The Caspian [Sea] Signed a letter of Mutual Understanding of Collaboration")	Nov. 5, 2003	150	0						
'Daryaa-ye Khazar Faghat baraa-ye 100 Saal Markaz-e Tavajjoh va Alaaghe-ye Afkaar-e Omumi Khaahad Bud' ("The Caspian Sea Will Only Be at the Center of Public Attention and Interest for 100 years")		405	3				3		
Aftaab-e Yazd ("Yazd's Sunshine"):	Nov. 6, 2003								
'Abutoraab Khosravi: Djamaalzaadeh Zabaan-e Kukhe va Baazar raa be Adabiyyaat-e Daastaani Tabdil Kard' ("Abutorab Khosravi: Djamalzadeh Transformed the Colloquial Language to Literary Fiction")		325	7		1		4		2
Totals		880	10	0	1	0	7	0	2

Newspaper Articles: World News									
On-line IRNA (Iranian News Agency):									
'Yek Ma'aareez-e Araaghi az Taghiir-e Barnaame-ye Jang-e Amrika Alaihe Araagh Sokhan Goft' ("An Iraqi Opponent of The American Change of Plans for War Against Iraq Spoke")	24-Mar-03	205	2		1		1		
Keyhaan ("Universe")	19-Mar-03								
'Towte'e-ye Kharaabkari-e Emrikaa-i-haa dar Kubaa Naakaam Mand' ("The American Plot for Destruction in Cuba Remained Unsuccessful")		123	4		2		2		
'Fard-e Mosallah-e Yamani 2 Karshenaas-e Emrikaai va Kaanaadai raa be Halaakat Resaand' ("Armed Yemeni Killed two American and Canadian Experts")		101	2				2		
'Tashdid-e Tadaabir-e Aminyati Atraaf-e Markaz-e Siyaasi-Nezaami-e Emrikaa dar Daakhel va Khaarej-e Keshvar' ("Repetition of Security Policy Surrounding the Internal and External American Political-Military Center")		264	2				1	1	
'E'teraaz-haa-ye Jahaani Alay-he Jang Sheddad Gereft' ("World Objections to Against War Intensified")		322	4				1	3	
'Nyoyork Zir-e Kontorol-e Niruhaaye Zedd-e-shoresh Gharaar Gereft' ("New York Was Placed Under the Control of Anti-riot Forces")		192	3			1	1	1	
'Saddam va Raamsfeld 18 saal ba'd' ("Saddam and Rumsfeld, 18 Years Later")		113	2		1			1	

'Emrika: Agar Saddam Ham Beravad Hamleh Mikonim' ("America: We Will Attack, Even if Saddam Leaves")		169	11						
				3	3		5		
'Lomond: Behbud-e Chehreh-ye Amrikaa dar Jahaan Gheyr-e-momken Ast' ("Le Monde: Improving America's Reputation in the World Is Impossible")		121	1						
				1					
'Dokhtar-e Saalvaador Alandeh Re'is-e Majles-e Nemaayandegaan-e Shili Shod' ("Salvador Allende's Daughter Elected as Speaker of Chile's House")		177	1						
				1					
'Vakil-e Mohjebbeh-ye Faraansavi raa be Maraasem-e Tahlif-e Vokalaa Raah Nadaadand' ("They Did Not Allow Covered [with Islamic Hejab] French Lawyer to the Lawyers' Oath Ceremony")		104	3						
				2					1
'Az Tars-e Marg Mordand' ("They Died of Fear of Death")		130	3						
				1			2		
'Hemaaghat-e Digar-e Bush' ("Bush's Other Foolishness")		324	5						
				1	1	1	2		
Hamshahri ("Fellow Townsman").	19-Mar-03								
'Mozaakeraat-e Iran va Yaman dar baareh-ye Bohraan-e Araagh' ("Discussions of Iran and Yemen about the Iraqi Crisis").		360	8						
				1	1	2	3		1
'Baaztaab-e Jahaani-e oltimaatom-e Bush be Saddam' ("World's Reaction to Bush's Ultimatum to Saddam")		615	18						
				5	4	2	4	1	2
'Sherkat-haa-ye Nafti-e Jahaan Kharid-e Naft-e Araagh ra Motafaghef Kardand' ("Oil Companies in the World Have Stopped the Purchase of Iraqi Oil")		252	5						
				3		1	1		

'Tasvib-e Ghanoon-e Ekhtihaar-aat-e Nakhost Vazir-e Felestin' ("Ratification of the Authority Law of the Palestinian Prime Minister")		546	9						
				7	1		1		
'Dar baareh-ye Esteghraar-e Niruhaa-ye Emrikaai Parleman-e Torkiyeh baar-e digar Tasmim Migirad' ("The Turkish Parliament once again Makes a Decision about the Settlement of the American Forces")		152	1						
				1					
'Kashf-e Selaah va Mavaadd-e Monfajereh dar Riaaz' ("Discovery of Weapons and Explosive Materials in Riyadh")		73	0						
Totals		4343	84	30	11	9	29	1	4
Children's Stories	1953 (no Day/Month listed)								
'Olaagh va Baar-e Namak' ("Donkey and The Load of Salt")		360	8						
				4		1	1	2	
'Gorg va Barreh' ("Wolf and Lamb")		380	13						
				5	2	1	4		1
'Khar dar Lebaas-e Shir' ("Donkey in Lion's Clothes")		323	10						
				4	3	1	1	1	
'Shekam-e Ghurbaaghe' ("Frog's Stomach")		256	11						
				1	4	3	3		
'Murcheh va Parvaaneh' ("Ant and Butterfly")		360	6						
						1	4	1	
Totals		1679	48	14	9	7	13	4	1
Speech Corpus:									
Linguistic Data Consortium (LDC): "CallFriend," an audio file of a telephone conversation between two friends.	July, 1997	2,266	37						
				8	11	5	5	5	3
Grand Totals		17,612	407	119	57	45	157	14	15

Appendix B

Short stories

	Word Count	<i>ra</i> Count	In Focus	Activated	Familiar	Uniquely Identifiable	Referential	Type Identifiable	Not Coded
Prologue	460	6	2	1	1		1		1
Introduction	576	11	3	2		5			1
Short Story: 'Sweet Hat' ("širin kola")	2505	89	30	17	9	32		1	
Short Story: 'Victorious Roman' ("fæth-e rumi")	2005	61	16	5	5	33			2
Totals:	5,546	167	51	25	15	70	1	1	4

Table 32 'Short Stories' Results Summary

Newspaper Articles: 'Sports' Section

	Word Count	<i>ra</i> Count
'Nekunaam: Mikhaahim Baa bacheh-haa-ye Paas be Shomaal Beravim' ("Nekunam: We Want to Go Up North with the <i>Pas</i> Players")	174	4
'Payaam-e Peykaan ham Ta'm-e Shekast ra Cheshid' (" <i>Payam-e Peykan</i> Also Tasted the Flavor of Defeat")	82	2
'Tahavvol dar Fa'aaliyat-haa-ye Varzeshi-e Mantagheh' ("Change in Sports' Activities in the Region")	122	1
'Perspolis Choob-e Kharid-haa-yash raa khord' (" <i>Persepolis</i> Was Punished Due to its Purchases [of new players]")	950	23
'Fulaad be Bazikonaanash Eidi Daad' (" <i>Fulad</i> " Gave its Players [New Year] Gifts")	133	0
'Aaghaazi Bad, Paayaani Badtar' ("A Bad Beginning, A Worse Ending")	943	21
'500 Hezar Dollar baraa-ye Seh Maah' ("500 Dollars for Three Months")	265	7
'Estili: Faghat Pahlevaan-e Zende h raa Eshgh Ast' ("Estili: <i>Hurray</i> Only for the Alive Hero")	229	4
Totals	2,898	62

Table 33 Eight Articles from the Sports Section of 'Iran' Newspaper

Newspaper Articles: ‘Culture’ Section

	Date Published	Word Count	ra Count
On-line <i>IRNA</i> (Iranian News Agency):	Nov. 5, 2003		
‘Daaneshgaah-haa-ye Keshvar-haa-ye Saheli-e Khazar Tafaahom-naameh-ye Hamkaari Emzaa Kardand’ (“The Universities of the Countries Bordering The Caspian [Sea] Signed a letter of Mutual Understanding of Collaboration”)		150	0
‘Daryaa-ye Khazar Faghat baraa-ye 100 Saal Markaz-e Tavajjoh va Alaaghe-ye Afkaar-e Omumi Khaahad Bud’ (“The Caspian Sea Will Only Be at the Center of Public Attention and Interest for 100 years”)		405	3
<i>Aftaab-e Yazd</i> (“Yazd’s Sunshine”):	Nov. 6, 2003		
‘Abutoraab Khosravi: Djamaalzaadeh Zabaan-e Kukhe va Baazar raa be Adabiyyaat-e Daastaani Tabdil Kard’ (“Abutorab Khosravi: Djamalzadeh Transformed the Colloquial Language to Literary Fiction”)		325	7
	Totals	880	10

Table 34 Three Articles from ‘Culture’ Sections of IRNA and Aftab-e Yazd

Newspaper Articles: ‘World’ Section

	Word Count	ra Count
‘Yek Ma’aarez-e Araaghi az Taghiir-e Barnaame-ye Jang-e Amrika Alaihe Araagh Sokhan Goft’ (“An Iraqi Opponent of The American Change of Plans for War Against Iraq Spoke”)	205	2

Table 35 One Article from 'World' Section of IRNA

	Word Count	ra Count
'Towte'e-ye Kharaabkari-e Emrikaa-i-haa dar Kubaa Naakaam Mand' ("The American Plot for Destruction in Cuba Remained Unsuccessful")	123	4
'Fard-e Mosallah-e Yamani 2 Karshenaas-e Emrikaai va Kaanaadai raa be Halaakat Resaand' ("Armed Yemeni Killed two American and Canadian Experts")	101	2
'Tashdid-e Tadaabir-e Aminyati Atraaf-e Markaz-e Siyaasi-Nezaami-e Emrikaa dar Daakhel va Khaarej-e Keshvar' ("Repetition of Security Policy Surrounding the Internal and External American Political-Military Center")	264	2
'E'teraaz-haa-ye Jahaani Alay-he Jang Sheddatt Gereft' ("World Objections to Against War Intensified")	322	4
'Neyoyork Zir-e Kontorol-e Niruhaaye Zedd-e-shoresh Gharaar Gereft' ("New York Was Placed Under the Control of Anti-riot Forces")	192	3
'Saddam va Raamsfeld 18 saal ba'd' ("Saddam and Rumsfeld, 18 Years Later")	113	2
'Emrika: Agar Saddam Ham Beravad Hamleh Mikonim' ("America: We Will Attack, Even if Saddam Leaves")	169	11
'Lomond: Behbud-e Chehreh-ye Amrikaa dar Jahaan Gheyr-e-momken Ast' ("Le Monde: Improving America's Reputation in the World Is Impossible")	121	1
'Dokhtar-e Saalvaador Alandeh Re'is-e Majles-e Nemaayandegaan-e Shili Shod' ("Salvador Allende's Daughter Elected as Speaker of Chile's House")	177	1
'Vakil-e Mohjebbeh-ye Faraansavi raa be Maraasem-e Tahlif-e Vokalaa Raah Nadaadand' ("They Did Not Allow Covered [with Islamic Hejab] French Lawyer to the Lawyers' Oath Ceremony")	104	3
'Az Tars-e Marg Mordand' ("They Died of Fear of Death")	130	3
'Hemaaghat-e Digar-e Bush' ("Bush's Other Foolishness")	324	5
Totals	2,140	41

Table 36 Twelve Articles from 'World' Section in *Keyhan*

	Word Count	ra Count
'Mozaakeraat-e Iran va Yaman dar baareh-ye Bohraan-e Araagh' ("Discussions of Iran and Yemen about the Iraqi Crisis").	360	8
'Baaztaab-e Jahaani-e oltimaatom-e Bush be Saddam' ("World's Reaction to Bush's Ultimatum to Saddam")	615	18
'Sherkat-haa-ye Nafti-e Jahaan Kharid-e Naft-e Araagh ra Motafaghef Kardand' ("Oil Companies in the World Have Stopped the Purchase of Iraqi Oil")	252	5
'Tasvib-e Ghanoon-e Ekhtiyaar-aat-e Nakhost Vazir-e Felestin' ("Ratification of the Authority Law of the Palestinian Prime Minister")	546	10
'Dar baareh-ye Esteghraar-e Niruhaa-ye Emrikaai Parleman-e Torkiyeh baar-e digar Tasmim Migirad' ("The Turkish Parliament once again Makes a Decision about the Settlement of the American Forces")	152	1
'Kashf-e Selaah va Mavaadd-e Monfajereh dar Riaaz' ("Discovery of Weapons and Explosive Materials in Riyadh")	73	0
Totals	1,998	42

Table 37 Six Articles from 'World' Section in *Hamshehri*

Children's Stories

	Word Count	ra Count	In Focus	Activated	Familiar	Uniquely Identifiable	Referential	Type Identifiable	Not Coded
'Olaagh va Baar-e Namak' ("Donkey and The Load of Salt")	360	8	4		1	1	1	1	
'Gorg va Barreh' ("Wolf and Lamb")	380	13	5	2	1	4			1
'Khar dar Lebaas-e Shir' ("Donkey in Lion's Clothes")	323	10	4	3	1	1	1		
'Shekam-e Ghurbaaghe' ("Frog's Stomach")	256	11	1	4	3	3			
'Murcheh va Parvaaneh' ("Ant and Butterfly")	360	6			1	4		1	
Totals	1,679	48	14	9	7	13	2	2	1

Speech Corpus

total word count: 2,266 words

37 occurrence of determiner phrases containing *ra*.

Appendix C

ra -marked DP found in Corpus	Phrase in which ra -marked DP occurred	Syntactic category	Highest Cognitive Status	Optional or Required
1. in o this ra	in o mā zabt ne-mi-kon-im. this- ra we record NEG-DUR-do-1PL 'We are not recording this'	Complement	Activated	Required * in-Ø mā zabt ne-mi-kon-im There are no contexts in which one could leave ra out as long as the demonstrative pronoun <i>in</i> ('this') is present.
2. in sohbat-e mā ro this conversation- ASP our ra	in-e ke in sohbat-e mā ro zabt mi-kon-an This-is that this conversation-LNKPT our-ra record DUR-do-3PL 'That's why they are recording this conversation of ours'	Complement	Activated	Required * in sohbat-e mā zabt mi-kon-an There are no contexts in which one could leave ra out as long as the demonstrative pronoun <i>in</i> ('that') is present.
3 & 4. (two instances of it) khodā ra God ra	A: Farah jān khub bud? Farah dear well was 'Was Farah well?' B: khub bud alhamdolelāh khodā ra shokr well was thank God (Arabic) God (Persian) ra thank (Arabic) '(She) was well, thank God.. Thanks be to God' A: khob. khodā ra shokr good God ra thank 'Good. Thanks be to God'	Complement	at least Activated	Required in any context * khodā shokr
5. Farah jān o Farah dear ra	Farah jān o bad hav ā kard-i umad-i Farah dear-ra bad climate did-2SG came-2SG 'You made dear Farah upset and came back'	Complement	Familiar	Required * Farah jān bad hav ā kard-i
6.	A: ham-in-tor mehmuni-ye. bale. in-jā	Complement	In Focus	Required

<p>jā-ye mā-ro place-LNKPT 1PL- ra</p>	<p>also-this-way party-is yes this-place</p> <p>mehmuni-ye o jā-ye shomā-m kheili khāli-ye party-is and place-LNKPT you-also very empty-is 'There are parties, one after another. Yes. Here there are parties, and your place is very empty [i.e., you are missed very much].'</p> <p>B: qorbān-e shomā. jā-ye mā-ro kheili khāli be-kon-in offering-LNKPT you place-LNKPT 1PL-ra very empty IMP-do-3PL '''Thank you'' [for missing us]. Keep an open spot for us.'</p>			<p>* jā-ye mā kheili khāli be- kon-in</p>
<p>7. dovvom-i ra second-i ra</p>	<p>dovvom-i ra ne-mi-khā-i be-zā-i? second-i ra NEG-DUR-want-2SG SUBJ-birth-2SG 'Do you not want to give birth to a second one[child]?'</p>	_____	_____	_____
<p>8 & 9 dovvom-i-r-am second-i-ra-also</p> <p>do tā-i-ro two CL-i-ra</p>	<p>go-ft-am shāyad dovvom-i-r-am</p> <p>say-PAST-1SG perhaps second-i-ra-also</p> <p>be-khā-i be-zā-i, ye dafe</p> <p>SUBJ-want-2SG SUBJ-birth-2SG one time</p> <p>do tā-i-ro bā ham bozorg kon-i two CL-i-ra with alsobig do-2SG 'I thought that perhaps you might want to give birth to a second one also, raising both [children] at the same time.'</p>	_____	_____	_____

10. har-chi - ra	na, man ke ⁷⁵ har-chi ra hesāb mi-kon-am no 1SG ‘as for me’ any-thing ra calculate DUR-do-1SG mi-bin-am un-jā gerun-tar-e DUR-see-1SG that-place expensive-COMP-is ‘No. As for me [or, “for my part”] whatever (i.e., goods) I calculate, I see it’s more expensive there...’	Adjunct	At least activated	Required * man ke har-chi hesāb mi- kon-am (possible to exclude ra , but meaning changes)
11. kheili chiz-ā-i ro many thing-PL-i ra	kheili chiz-ā-i -ro ke lāzem many thing-PL-i-ra that necessary dār-am az in-jā mi-khar-am have-1SG from this-place DUR-buy- 1SG ‘many of the things I need I buy from there.’	_____	_____	_____
12. khabar-ā ro news-PL ra	khob, mā juyā-ye hāl-et bud-im. good 1PL seek-LNKPT health-2SG was-1PL ‘Well, we had wondered how you were.’ az Iran soraq-e -to dāsht-im. from Iran seek-LNKPT-2SG have-PAST-1PL bar-gashtan-am soraq-e-to dāsht-am. back-turn[INFINITIVE]-also seek-LNKPT-2SG have-PAST-1SG ‘Coming back, we’d asked about you as well.’ bā Shamsi khānum umad-in, khabar-ā-ro with Shamsi Ms. come-PAST-2PL news-PL-ra	Complement	At least activated	Required * khabar-ā āvord-an. ra is obligatory with PL marker, unless we mean news in the general sense, as in “oh, what news they brought with ‘em!” (‘khabarhā āvordan!’) Also, ‘khabar āvord-an’ would be more appropriate in the case where Shamsi brings the news; once again, a change in meaning.

⁷⁵ ‘ke’ is a Relative marker, but here, it is used idiomatically to mean “for my part”, or “as for me”.

	āvord-an. bring-PAST-3PL			
13. chamedun in-ā ro luggage this-PL ra	chamedun in-ā -ro gozāsht-im manzel-e Feri jān luggage this-PL-ra put-PAST-1PL residence- LNKPT Feri dear 'We left the luggage "and the such" at dear Feri's place.'	Complement	Uniquely identifiable	Required * chamedun in-ā gozāsht-im manzel-e Feri jān
14. un taraf-ā ro that around-PL ra	bad raft-im un taraf-ā-ro gasht-im next go-PAST-1PL that around-PL-ra visit-1PL 'then we went sight-seeing around there.'	Adjunct	Uniquely identifiable	Required (in this context) # raft-im un taraf-ā gasht-im My native judgment tells me that without ra , there is somewhat of a change in meaning. The presence of ra seems to put more emphasis on the "un tarafa": 'raft-im un taraf-ā gasht-im' the emphasis is on "gashtim" 'raft-im un taraf-ā-ro gasht- im' The emphasis is on 'un tarafa' Once again, there is a change in meaning.
15. hamdiga-r-am	ettefāqan un-jā chandin bār hamdiga-r- am ⁷⁶ dide bud-an	Complement	At least activated	Required

⁷⁶ It is not entirely clear if the 'r' is from 'hamdigar' ('each other') or from a '**ra**'. In careful speech we get 'hamdigar ra hæm' ([r] [ra] [h] → [r])
A natural question that may follow would be 'how do we know there is a **ra** in fast speech?' An explanation may be that **ra** is there because
'each other' refers to Ali and Feri, who are at least Activated. The claim is that anything that is Activated or higher must co-occur with **ra**.
See section 4.2.1.1.3

each other- ra -also	coincidentally that-place several times each other-ra-also see was-3PL 'coincidentally, they had seen each other there several times.'			* tefāqan un-jā handin bār hamdigar am (as opposed to ' hamdigar ra ham ') dide bud-an
16. Shamsi khānum- o -am Shamsi Ms. -ra -also	Shamsi khānum- ⁷⁷ o -am dide bud-an. Shamsi Ms. -ra -also see was-3PL 'They had seen Ms. Shamsi as well', ⁷⁸	Complement	Activated	Required to preserve meaning * Shamsi khānum [h]am dide bud-an.
17. Shamsi khānum o Shamsi Ms. ra	man Shamsi khānum-o na-tunest-am be- bin-am, 1SG Shamsi Ms.-ra NEG-able-PAST-1SG SUBJ-see-1SG 'I was unable to see Ms. Shamsi'	Complement	At least activated	Required * man Shamsi khānum na- tunest-am be-bin-am
18. kār-am o work-1SG ra	man ye māh-e kār-am o 1SG one month-LNKPT work-1SG ra shoro kard-am start do-PAST-1SG '[it's only been] one month [since] I began my work.'	Complement	In Focus	Required *man ye mah-e kar-am shoro kard-am
19. nesf-esh o half-3SG ra	nesf-esh-o tu khāb pāk kar... [hesitation: 'kardam'] half-3SG-ra in sleep clean do... '[I] cleaned half of it [as if] in sleep'	Adjunct	In Focus	Optionality: Required * nesf-esh tu khab pak kar[dam]
20. man o 1SG ra	be amir goft-am " man-o mi-bar-i to Amir say-PAST 1SG-ra DUR-take- 2SG	Complement	At least activated	Required * man mi-bari mahi be-khar-

⁷⁷ **ra** here, too, is contracted. See section 4.2.1.1.3

⁷⁸ Note that this sentence in isolation is ambiguous in that it can be interpreted as 'Ms. Shamsi had seen them'; however, the context dictates that 'Ms. Shamsi' here is in direct object position. Indeed, 'Ms. Shamsi was in object position, there would be no need for **ra**. However, native judgment tells us otherwise. See discussion in section 4.2.1.1.3.

	māhi be-khar-am” fish SUBJ-buy-1sg ‘I asked Amir “would you take me to buy fish” ’			am
21 man o 1SG ra	umād man-o bord come-PAST-3SG 1SG-ra take-PAST ‘s/he came [and] took me’	Complement	At least activated	Required * man mi-bari mahi be-khar-am
22 & 23 nesf-esh-o half-2SG-ra	nesf-esh-o qāch zad, nesf-esh-o na-zad half-2SG-ra cuthit-3SG half-2SG-ra NEG-hit-3SG ‘he cut half of it, [but] did not cut the other half.’	Complement	At least activated	(22): Required * nesf-esh qach zad (23): Required * nesf-esh qach na-zad
24 kesāfat kāri-ā-sh o dirty work-PL-3SG ra	majbur bud-am in kesāfat kār-i-ā-sh-o forced was-1SG this dirty work-i-PL-3SG-ra khod-am be-kon-am self-1SG IMP(?) -do-1SG ‘I was forced to do the dirty work myself.’	Complement	Activated	Required * kesāfat kāri-ā-sh khod-am be-kon-am
25 amir jun o Amir dear ra	dige amir jun-o ham-in-tor, amin jān, imān jān else Amir dear-ra also-this-way Amin dear Iman dear ‘also, likewise [greet] dear Amir, dear Amin, dear Iman’	Complement	Familiar	Required * amir jun hamintor
26& 27 showhar-e chiz o husband-LNKPT thing ra mehrak o Mehrak ra	rasti, showhar-e chiz-o did-i, mehrak-o ^{79?} by the way husband-LNKPT thing-ra see-2SG Mehrak-ra ‘By the way, did you see so-and-so’s husband, Mehrak’s [husband]?’	Complement	Familiar	(26): Required * showhare chiz didi (27): Required * [showhare] Mehrak
27	Mehrak ⁸⁰ o?	Complement	In Focus	Required

⁷⁹ Note: object head drop; reference is being made to Mehrak’s husband

Mehrak ra Mehrak ra	Mehrak ra 'Mehrak's [husband]?			* Mehrak [didi]?
28 liāqat-esh o worthy-3SG ra	enshālā ke liāqat-esh-o dāsht-e bāsh-e God willing that(CONJ) worthy-3SG-ra have-PAST- 3SG-LNKPT be-2SG 'God willing, may he have her worth [i.e., be worthy of her]'	Complement	Activated	Required * liāqat-esh dāsht-e bāsh-e
29 jelo-e khod-esh o front-of self-3SG ra	age, bale, jelo-e khod-esh-o na-gir-e mi- sh-e if yes front-of self-3SG-ra NEG-control-3SG DUR-become-3SG 'Yes, if she doesn't control herself, she will become [fat]'	Complement	At least familiar, if not In Focus	Required * jelo-e khod-esh na-gir-e
30 shekl-e khub-e un o face-LNKPT good-LNKPT 3SG ra	shekl-e khub-e un-o ne-mi-dun-am face-LNKPT good-LNKPT 3SG-ra NEG-DUR- know-1SG 'As for his good looks, I don't know.'	Adjunct	Familiar	Required * man shekl-e khub-e un ne- mi-dun-am
31 khod-esh- Ø -am self-3SG- ra -also	be khosus alān khod-esh-Ø ⁸¹ - am (i.e., contracted khod-e-sh ra ham ⁸²) to regard now self-3SG-ra-also lāqar kard-e skinny do-PSPT 'especially now [that] she has made herself skinnier [i.e., lost weight].'	Complement	In Focus	Required * be khosus alān khod-esh [h]am lāqar kard-e

⁸⁰ Note: once again, object head drop; reference is being made to Mehrak's husband

⁸¹ **ra** is contracted here.

⁸² See section 4.2.1.1.3

32 aks-ā ro picture-PL ra	aks-ā ro ferest-ād picture-PL ra sent-3SG 'He sent the pictures.'	Complement	In Focus	Required * aks-ā ferest-ād
33 senn-esh o age-3SG ra	jelo-ye mu-hā-sh-am ye khorde kam-e, front-LNKPT hair-PL-3SG-also one little less-is senn-esh o bishtar neshun mi-d-e age-3SG ra more show DUR-give- 3SG '[there is] fewer hair in front [and it] shows his age [to be] more [than it actually is]'	Complement	Uniquely Identifiable	Required to preserve meaning ⁸³ * senn-esh bishtar neshun mi- d-e
34 kas-i ro person-i ra	kas-i ro peidā na-k-ard-am. person-i ra find NEG-do-PAST-1SG 'I did not find anyone.'	_____	_____	_____

⁸³ Based on this example and others, I am toying with an idea that could potentially be explored in future versions of the *ra* paper. It appears to me that we could argue for a function of *ra* that has not been mentioned anywhere in the literature, as far as I know: *ra* as a morphological causative, adding a controlling participant that is causing an action. Observe:

- (a) senn-esh bishtar neshum mide
age-3rd-Sing. more shows
'his age shows more'
- (b) senn-esh o bishtar neshum mi
's/he shows his age to be more'

Note that in (a) 'age' (of the 3rd person singular) is the causee (the agent of the caused event), whereas in (b), the 3rd person singular is the agent. This example and examples like it needs further exploration/explanation.

35 man o 1SG ra	<p>valikan Sepehr ke vaqti aziyat be-kon-e but Sepehr when (REL ADV) time trouble SUBJ-do-3SG</p> <p>man o ke man ne-mi-tun-am be-d-am 1SG ra that(CONJ) I NEG-DUR-can-1SG SUBJ-give-1SG</p> <p>dast-e un yā dast-e shohar-e un. hand-LNKPT 3SG or hand-LNKPT husband-LNKPT 3SG</p> <p>'But when Sepehr troubles me, I can't [just] give him to her or to her husband.'</p>	Adjunct	Activated	Required * Sepehr ke vaqti aziyat be-kon-e man ke man me-mi-tun-am be-d-am dast-e un
36 arusi-sh o wedding-3SG ra	<p>mage in-ke amu Mehrang arusi-sh-o unless this-that uncle Mehrang wedding-3SG-ra</p> <p>ye... che-mi-dun-am, shesh māh ba'd one what-DUR-know-1SG six month later</p> <p>be-ndāz-e aqab, SUBJ-put-3SG postpone</p> <p>'Unless uncle Mehrang, I don't know, postpones his wedding for six months or so'</p>	Complement	At lease Uniquely Identifiable, but most likely Familiar	Required * ... amu Mehrang arusi-sh ...
37 un ye zarra-sh o 3SG one little-3SG ra	<p>un ye zarra-sh-o ādam in-var 3SG one little-3SG-ra person this side</p> <p>un-var mi-kon-e, bolan mi-sh-e mi-r-e that side DUR-do-3SG up DUR-</p>	Complement	Familiar	Required * un ye zarra-sh ādam ...

	become-3SG DUR-go-3sg ‘that little bit of it [i.e., time] one could “give or take” (and) pack up and go.’			
38 zarf-ā-sh o dishes-PL-3PL ra	dāsht-am zarf-ā-sh-o mi-shost-am had-1SG dishes-PL-3PL-ra DUR-clean-1SG ‘I was cleaning its dishes.’	Complement	Uniquely Identifiable	Required * un ye zarra-sh ādam ...
39 Peymān o Peyman ra	kheili Peymān-o salām be-re-sun-ā very Peyman-ra hello IMP-carry-2SG-EMPHATIC ‘Do give my highest regards to Peyman.’	Adjunct ⁸⁴	Familiar	Required * kheili Peymān salām be-re-sun-ā

⁸⁴ Note that “Peyman” here could be considered adjunct. See section 4.2.1, # 26.